BACHELOR OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING

A CNN-based Pipeline using Plane-wise Ensemble Technique for Classifying Alzheimer's Disease from 3D MRI Images

Mahtab Nur Fardin
190041112
Md. Irfanur Rahman Rafio
190041125
Md. Jubayer Islam
190041129

Department of Computer Science and Engineering
Islamic University of Technology
June, 2024

A CNN-based Pipeline using Plane-wise Ensemble Technique for Classifying Alzheimer's Disease from 3D MRI Images

Mahtab Nur Fardin
190041112
Md. Irfanur Rahman Rafio
190041125
Md. Jubayer Islam
190041129

Department of Computer Science and Engineering
Islamic University of Technology
June, 2024

Declaration of Candidate

This is to certify that the work presented in this thesis is the outcome of the analysis and experiments carried out by **Mahtab Nur Fardin**, **Md. Irfanur Rahman Rafio**, and **Md. Jubayer Islam** under the supervision of **Dr. Md. Hasanul Kabir**, Professor, Department of Computer Science and Engineering and co-supervision of **Sabbir Ahmed**, Assistant Professor, Department of Computer Science and Engineering, Islamic University of Technology, Dhaka, Bangladesh. It is also declared that neither this thesis nor any part of it has been submitted anywhere else for any degree or diploma. Information derived from the published and unpublished work of others have been acknowledged in the text and a list of references is given.

Dr. Md. Hasanul Kabir
Professor
Department of Computer Science and Engineering
Islamic University of Technology (IUT)
Date: June 04, 2024

Sabbir Ahmed
Assistant Professor
Department of Computer Science and Engineering
Islamic University of Technology (IUT)
Date: June 04, 2024

Mahtab Nur Fardin
Student ID: 190041112
Date: June 04, 2024

Md. Irfanur Rahman Rafio
Student ID: 190041125
Date: June 04, 2024

Md. Jubayer Islam
Student ID: 190041129
Date: June 04, 2024

Contents

1 Introduction **1**

1.1 Motivation and Scope . . . 3
1.2 Problem Statement . . . 5
1.3 Research Challenges . . . 5
1.3.1 Variability in Brain Morphology . . . 5
1.3.2 Multi-Class Classification Complexity . . . 6
1.3.3 Computational Cost . . . 6
1.4 Contribution . . . 8
1.5 Organization . . . 8

2 Related Works **10**

2.1 Conventional Approaches . . . 10
2.1.1 Binary Classification Using Machine Learning Techniques . . . 10
2.1.2 Image Processing Methods . . . 11
2.1.3 Limitations . . . 11
2.2 Deep Learning Based Approaches . . . 11
2.2.1 Dominance of CNN Models . . . 12
2.2.2 Architecture Overview . . . 12
2.2.3 Multi-class Classification . . . 20
2.2.4 Utilization of Single and Multi-modal Approaches . . . 21
2.2.5 Preprocessing Techniques . . . 23
2.3 Summary and Limitations of Existing Architectures . . . 25

3 Proposed Methodology **27**

3.1 Architecture Overview . . . 27
3.2 Model Architecture . . . 29
3.3 Model Integration . . . 29
3.4 Projector Functions . . . 30

3.4.1 Midplane Projector 31
3.4.2 Average Projector 32
3.4.3 Max Variance Projector 32
3.4.4 Variance Weighted Average Projector 33
3.4.5 Linear Learnable (LL) Projector 33

4 Results and Discussion **35**
4.1 Dataset 35
4.2 Experimental Setup 36
4.2.1 Data Preparation 36
4.2.2 Hyper-parameter Settings 38
4.3 Quantitative Evaluation 40
4.3.1 Evaluation Metrics 40
4.3.2 Baseline Models 43
4.3.3 Plane-wise Ensemble Models 47
4.3.4 Comparative Analysis and Discussion 52

5 Conclusion **58**

References **60**

List of Figures

2.1 A simple CNN architecture that extracts the features from the spatial information and then classify into different classes [33] 13
2.2 The architecture of AlzheimerNet, a fine-tuned InceptionV3 [34] . . . 13
2.3 Schematic representation of proposed ensemble model [21]. 14
2.4 Network architecture of 3MT [19]. 15
2.5 Details of the CNN transformer for encoding 3D images [19]. 16

3.1 Diagram comparing a regular ensemble model (c) with our proposed plane-wise ensemble technique (f) 28
3.2 Midplane Projections: Axial Plane 32
3.3 Average Projections: Axial Plane 32
3.4 Max Variance Projections: Axial Plane 33
3.5 Variance Weighted Average Projections: Axial Plane 33

4.1 Skull-stripping: The process of removing non-brain tissues from MRI images using semantic segmentation 38

List of Tables

2.1 Summary of Representative Works on Alzheimer's Disease Classification 26

4.1 Dataset Distribution 36
4.2 Data Split Distribution 39
4.3 Performance Analysis for AlexNet 44
4.4 Performance Analysis for VGG-16 44
4.5 Performance Analysis for ResNet-50 45
4.6 Performance Analysis for DenseNet-169 45
4.7 Performance Analysis for Vision Transformer B16 46
4.8 Performance Analysis for Traditional Ensemble Model (ResNet-50, AlexNet, VGG-16) 46
4.9 Performance Analysis for Triple ResNet (using Midplane Projector) . . 47
4.10 Performance Analysis for Triple DenseNet (using Midplane Projector) 48
4.11 Performance Analysis for Triple ResNet (using Average Projector) . . 48
4.12 Performance Analysis for Triple DenseNet (using Average Projector) . 49
4.13 Performance Analysis for Triple ResNet (using Max Variance Projector) 49
4.14 Performance Analysis for Triple DenseNet (using Max Variance Projector) 50
4.15 Performance Analysis for Triple ResNet (using Variance Weighted Average Projector) 50
4.16 Performance Analysis for Triple DenseNet (using Variance Weighted Average Projector) 51
4.17 Performance Analysis for Triple ResNet (using LL Projector) 51
4.18 Performance Analysis for Triple DenseNet (using LL Projector) 52
4.19 Model Performance Comparison 54
4.20 Confidence Intervals for Model Accuracy 56
4.21 Comparison with State-of-the-art Models 57

Abstract

Alzheimer's disease (AD) is a chronic neurodegenerative condition that progressively damages brain cells, resulting in memory and cognitive decline and eventually impeding basic functionalities. With over 55 million people worldwide affected by dementia, a number anticipated to rise significantly, the urgency for early diagnosis becomes paramount. While a definitive cure remains elusive, early intervention is crucial in mitigating disease progression and enhancing patient outcomes. This research investigates the potential of deep learning models for classifying Alzheimer's Disease, emphasizing the challenges in Mild Cognitive Impairment (MCI) classification, and introduces a CNN-based pipeline utilizing a plane-wise ensemble technique for 3D MRI image classification. To manage the complex nature of 3D MRI data, a CNN-based pipeline is proposed that makes use of a plane-wise ensemble technique. Decomposing the 3D image into axial, coronal, and sagittal planes, and using an ensemble of 2D CNN models trained on the axial, coronal, and sagittal planes, the system attempts to include multi-view data and improve classification accuracy. This methodology also leverages projector functions to map the 3D volumes into a series of 2D images and tackles the computational challenges presented by 3D data, resulting in a more efficient and practical process even with constrained computation resources.

Chapter 1

Introduction

Alzheimer's disease (AD) is a pervasive and devastating neurodegenerative disorder characterized by the gradual deterioration of brain tissues, leading to a decline in memory, cognitive functions, and ultimately, a loss of fundamental abilities. As the global population ages, the prevalence of dementia, including AD, continues to escalate, with over 55 million individuals affected worldwide and this figure is expected to rise to 78 million in 2030 and 139 million in 2050 [28].

Traditionally, biomarkers, focusing on critical brain regions like the hippocampus, parietal lobe, and amygdala, have been fundamental in identifying atrophy indicative of AD [4]. Recent studies, however, have showcased the high potential of deep learning models in the accurate classification of AD, prompting a paradigm shift in diagnostic methodologies [1].

Before delving into our core research, it is essential to comprehend the domain intricacies. Experts categorize potential AD patients into three classes, a departure from the conventional binary classification of AD and normal cognitive function (CN). The introduction of Mild Cognitive Impairment (MCI) allows for early detection, recognizing that most MCI patients progress to AD within 3 to 6 years [23]. MCI poses a unique challenge, as early-stage MCI resembles CN, while late-stage MCI bears similarity to AD.

Examining the modalities instrumental in diagnosing AD reveals a shifting landscape. Outdated methods like CT scans and EEG have given way to more sophisticated approaches. Magnetic Resonance Imaging (MRI), offering detailed structural images, has become the primary modality due to its widespread availability. Positron Emission Tomography (PET), while providing functional insights, is invasive, limiting its use. Diffusion Tensor Imaging (DTI), though yielding high-quality data, faces challenges

in terms of accessibility. Remarkably, in Alzheimer's classification, the integration of text data like age, gender, MMSE scores, and genetic information yields nuanced insights into disease dynamics, adding to the enhancement of accuracy. Demographic factors such as age and gender capture variations, while MMSE scores provide standardized cognitive assessments. Genetic data contributes to understanding hereditary patterns, collectively enhancing the precision of classification models for a comprehensive analysis of Alzheimer's disease [39].

In recent years, research centers have aggregated substantial medical and image data, sharing it publicly for the benefit of researchers engaged in Artificial Intelligence (AI) development for Alzheimer's Disease. These online datasets provide crucial biomarker information, including neuroimaging modalities, genetic data, and clinical and cognitive assessments. The most prominent datasets are:

- **Alzheimer's Disease Neuroimaging Initiative (ADNI) [15]:** ADNI, a longitudinal and multicenter study, serves as a prominent dataset. It includes ADNI-1, ADNI-GO, ADNI-2, and ADNI-3. ADNI aims to assess the progression of Mild Cognitive Impairment (MCI) and early AD, utilizing Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), biological markers, and clinical and neuropsychological assessments. ADNI datasets encompass various data types, such as clinical information, genetic data, MRI and PET images, and biospecimens.
- **Australian Imaging, Biomarker, & Lifestyle Flagship Study of Aging (AIBL) [10]:** AIBL compiles imaging and medical data from individuals with AD, those with MCI, and cognitive healthy individuals.
- **Open Access Series of Imaging Studies (OASIS)[22]:** OASIS, designed to share neuroimaging brain datasets, includes OASIS-1 with 434 MRI scans, OASIS-2 with 373 MRI scans, and OASIS-3 with 2,168 MRIs and 1,608 PET scans.
- **National Alzheimer's Coordinating Center (NACC)[5]:** NACC, established as a cornerstone in Alzheimer's research, is a pivotal dataset that makes invaluable contributions to our understanding of the disease. NACC not only provides essential data but also serves as a nexus for standardizing and harmonizing diverse datasets. Its comprehensive collection includes clinical, genetic, and neuroimaging data, fostering a holistic approach to Alzheimer's research.

Recent advancements in Alzheimer's disease classification have witnessed the effectiveness of deep learning models. Building upon that, this report presents a Convo-

lutional Neural Network (CNN)-based pipeline for classifying Alzheimer's Disease from 3D MRI images. The pipeline employs a plane-wise ensemble technique, introducing a strategic approach that capitalizes on specialized models for distinct imaging planes. This innovative technique aims to harness anatomical information during both model training and evaluation, demonstrating remarkable improvements in the classification accuracy of 3D MRI images.

1.1 Motivation and Scope

Early and accurate diagnosis of AD is crucial for effective patient management and the development of therapeutic strategies. Magnetic Resonance Imaging (MRI) has emerged as a powerful non-invasive tool for diagnosing AD due to its high-resolution imaging capabilities. However, manual analysis of 3D MRI images is time-consuming, subjective, and prone to errors, necessitating the development of automated and reliable diagnostic methods [17].

Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have shown remarkable success in image classification tasks [24]. CNNs have the potential to revolutionize the field of medical imaging by automating the analysis process, reducing human error, and enhancing diagnostic accuracy. Despite these advancements, the complexity of 3D MRI data presents unique challenges, including high dimensionality and computational intensity, which often limit the performance and applicability of standard CNN models.

To address these challenges, the motivation for this thesis is to explore a novel CNN-based pipeline that leverages a plane-wise ensemble technique for classifying Alzheimer's disease from 3D MRI images. By decomposing the 3D MRI data into 2D planes and employing an ensemble of CNN models trained on these planes, the proposed method aims to improve classification accuracy while managing computational complexity. This approach not only capitalizes on the strengths of 2D CNNs but also integrates the multi-view information inherent in 3D data, offering a promising solution for robust AD diagnosis.

The significance of this research lies in its potential to enhance the early detection of Alzheimer's disease, thereby enabling timely intervention and improved patient outcomes. Furthermore, by optimizing the computational efficiency of the diagnostic process, this approach can facilitate broader clinical adoption and integration into existing diagnostic workflows.

Recent research has demonstrated the efficacy of CNNs in medical imaging, particularly in the classification of neurological conditions from MRI data. Different studies [20], [21], [29], [38] have shown that deep learning models can achieve high accuracy in distinguishing between AD and healthy controls. Additionally, works by Payan and Montana have explored 3D CNNs for AD diagnosis [27], highlighting the potential of deep learning in this domain. However, these studies often grapple with the computational demands and complexity associated with processing 3D MRI data.

Despite the promising results, existing research is limited by several factors. The high dimensionality of 3D MRI data leads to significant computational requirements, making the training and deployment of 3D CNN models resource-intensive. Additionally, many current approaches also overlook the potential benefits of integrating multi-view information from different planes of 3D MRI data, which can enhance diagnostic accuracy.

This thesis aims to bridge these gaps by developing, implementing, and evaluating a CNN-based pipeline designed for the classification of Alzheimer's disease using 3D MRI images. The key aspects covered include:

1. **Data Preprocessing**: Detailed examination of preprocessing techniques to standardize 3D MRI data, including normalization, resizing, and slice extraction, ensuring compatibility with the proposed CNN models.

2. **Model Architecture**: Design and implementation of CNN architectures tailored for 2D plane classification, followed by an ensemble approach that integrates predictions from multiple planes (axial, coronal, and sagittal).

3. **Training and Validation**: Strategies for effectively training the CNN models, including data augmentation, cross-validation, and optimization of hyperparameters to enhance model performance and generalizability.

4. **Performance Evaluation**: Comprehensive evaluation of the pipeline using established metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). Comparisons with existing methods to highlight improvements and potential benefits.

By addressing these components, the thesis aims to contribute to the field of medical imaging and Alzheimer's disease diagnosis, providing a scalable and efficient solution for early detection and improving the overall quality of patient care. The proposed plane-wise ensemble technique not only enhances diagnostic accuracy but also significantly reduces the computational burden, making the model feasible for clinical

application even with limited computational resources. This innovation has the potential to make advanced diagnostic tools accessible to a wider range of healthcare settings, thereby improving patient outcomes on a global scale.

1.2 Problem Statement

Develop a model which, given a 3D MRI image, can classify it into one of the following classes:

- Alzheimer's Disease (AD)
- Cognitive Normal (CN)
- Mild Cognitive Impairment (MCI)

Each 3D MRI image presents a complex and high-dimensional dataset that requires advanced techniques for accurate classification. Traditional manual analysis methods are labor-intensive and susceptible to subjective bias and errors. Therefore, there is a pressing need for automated, reliable, and efficient diagnostic tools to support the early detection and differentiation between these categories.

The goal is to provide a robust and scalable tool for aiding the diagnosis and monitoring of Alzheimer's Disease and related conditions. This will ultimately improve patient outcomes and support clinical decision-making.

1.3 Research Challenges

The development of a CNN-based pipeline for classifying Alzheimer's disease from 3D MRI images involves several significant challenges. These challenges span computational costs, variability in brain morphology, and the complexity of multi-class classification. Addressing these challenges is crucial for the successful implementation and accuracy of the proposed diagnostic tool.

1.3.1 Variability in Brain Morphology

Detecting brain atrophies associated with Alzheimer's disease is complicated by the natural variability in brain structure. Factors contributing to this variability include:

- **Gender Differences**: Male and female brains exhibit structural differences, which can affect the model's ability to generalize across genders.

- **Age-related Changes**: The brain undergoes significant changes over a person's lifespan, adding another layer of complexity to the classification task. Age-related atrophies might be mistaken for disease-specific patterns.
- **Demographic and Genetic Diversity**: Variations in brain structure can also be attributed to demographic factors (e.g., ethnicity, lifestyle) and genetic predispositions, complicating the detection of AD-related changes.

Developing a model that can account for these variabilities requires a robust training dataset that adequately represents these diverse factors. It also necessitates sophisticated preprocessing and augmentation techniques to ensure the model is exposed to a wide range of anatomical variations.

1.3.2 Multi-Class Classification Complexity

Classifying 3D MRI images into three categories (AD, CN, MCI) introduces additional complexity compared to binary classification. Specific challenges include:

- **Interclass Similarity**: The Mild Cognitive Impairment (MCI) class often exhibits characteristics that overlap with both Alzheimer's Disease (AD) and Cognitive Normal (CN) classes, making it difficult for the model to distinguish between these states accurately.
- **Imbalanced Data**: The prevalence of AD, CN, and MCI in available datasets may not be evenly distributed, potentially leading to biased model performance if not adequately addressed.
- **Diagnostic Ambiguity**: The progression from CN to MCI to AD is a continuum, and the boundaries between these classes are not always clear-cut. This ambiguity can lead to misclassification and reduced model accuracy.

1.3.3 Computational Cost

One of the primary challenges in developing a deep learning model for 3D MRI image classification is the high computational cost associated with training and inference. Key issues include:

- **High Dimensionality of Data**: 3D MRI images contain a vast amount of data, leading to high memory and processing requirements. The "Curse of Dimensionality" exacerbates this issue, as the number of parameters in the model increases exponentially with the dimensionality of the data. This makes learning

more difficult and necessitates large amounts of training data to avoid overfitting.

- **Model Complexity**: Standard 3D CNN models are computationally expensive due to the large number of parameters and layers required to capture spatial features in three dimensions. High-dimensional data also complicates optimization, making it more susceptible to local minima and increasing the difficulty of finding a global optimum. Additionally, outlier analysis becomes less meaningful in high-dimensional spaces.
- **Hardware Limitations**: Many research institutions and healthcare facilities may not have access to high-performance computing resources, limiting the feasibility of training complex 3D CNN models. The significant memory and processing power required to handle 3D MRI data can strain available hardware, leading to slower training times and reduced model performance.

To mitigate these issues, the proposed plane-wise ensemble technique aims to decompose 3D MRI data into 2D planes, significantly reducing computational demands and making the training process more efficient. This approach reduces the number of parameters, simplifies optimization, and lowers the data requirements, making the model more feasible for training on available hardware. Additionally, by focusing on 2D planes, the method can leverage the strengths of 2D CNNs, which are less computationally intensive than their 3D counterparts.

To address these challenges, the research will focus on:

- **Balanced Sampling**: Techniques to ensure that the dataset is balanced by sampling from each class multiple times, ensuring statistical significance and robustness of the method. Strong measures are taken to prevent data leakage, maintaining the integrity of the training process.
- **Ensemble Techniques**: Using multiple CNNs trained on different 2D planes to capture more nuanced features and improve overall classification robustness. This approach leverages the strengths of 2D CNNs while integrating multi-view information from 3D data.
- **Enhanced Model Training**: Employing strategies such as cross-validation and hyperparameter tuning to optimize model performance and mitigate the effects of interclass similarity. These techniques help in fine-tuning the model parameters and ensuring that the model generalizes well across different subsets of the data.

By tackling these computational, morphological, and classification challenges, this research aims to develop a more accurate, efficient, and generalizable CNN-based diagnostic tool for Alzheimer's disease and related conditions.

1.4 Contribution

This thesis presents several key contributions to the field of medical imaging and Alzheimer's disease diagnosis:

- **Introduced a novel plane-wise ensemble technique** that addresses the computational challenges of 3D MRI data by decomposing it into 2D planes (axial, coronal, and sagittal) and employing an ensemble of CNN models trained on these planes. This method reduces computational complexity while leveraging the strengths of 2D CNNs, enhancing classification accuracy and efficiency.
- **Developed projector functions** that map 3D MRI images into 2D inputs for the CNN models. These functions transform high-dimensional 3D data into a manageable format for 2D CNNs, maintaining essential structural information and simplifying the data processing.
- **Benchmarked the performance** of different CNN models, systematically evaluating and comparing the models. This benchmarking provides insights into the most effective models and plane orientations for the classification task, ensuring the use of the most accurate and reliable configurations for Alzheimer's disease diagnosis.

1.5 Organization

The thesis is structured to provide a comprehensive exploration of the Plane-wise Ensemble Technique for classifying Alzheimer's Disease from 3D MRI images. The organization ensures a logical flow of information, guiding the reader from the introduction to the conclusion while aligning with the research objectives.

Chapter 2 conducts a thorough review of existing literature on Alzheimer's Disease diagnosis and medical image classification techniques. It discusses the limitations of traditional methods and highlights the potential benefits of utilizing CNN-based approaches, laying the theoretical groundwork for the Plane-wise Ensemble Technique and its relevance in the field of medical imaging.

Chapter 3 presents an overview of the Plane-wise Ensemble Approach, detailing the decomposition of 3D MRI data into axial, coronal, and sagittal planes. It explains the rationale behind leveraging anatomical information from different imaging planes to enhance classification accuracy and introduces the workflow and integration of specialized models within the ensemble framework.

Chapter 4 details the implementation of the CNN-based pipeline for classifying Alzheimer's Disease using the Plane-wise Ensemble Technique. It discusses the experimental setup, including dataset selection, model training, and evaluation metrics, and analyzes the results to showcase the advancements in classification accuracy achieved through the proposed approach.

Chapter 5 summarizes the key findings of the research, emphasizing the contributions of the Plane-wise Ensemble Technique. It discusses the implications of the findings in relation to the research objectives and existing knowledge in the field, acknowledges study limitations, and suggests future research directions to build upon the current work, providing a comprehensive wrap-up of the thesis.

Chapter 2

Related Works

2.1 Conventional Approaches

Earlier methods primarily focused on binary classification and traditional image processing techniques, which laid the groundwork for more sophisticated models.

2.1.1 Binary Classification Using Machine Learning Techniques

One of the initial strategies for Alzheimer's disease diagnosis involved binary classification, distinguishing between Alzheimer's disease (AD) and Cognitive normal (CN) individuals. Machine learning techniques such as Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), and Principal Component Analysis (PCA) [2], [20], [41] were commonly employed in these efforts. SVMs, known for their effectiveness in high-dimensional spaces, were utilized to create hyperplanes that best separated the AD and CN classes based on features extracted from neuroimaging data. LDA, on the other hand, aimed to find the linear combinations of features that best separated the two classes. PCA was often used to reduce the dimensionality of the data, retaining the most informative components for subsequent classification tasks. These machine learning approaches provided a foundation for early diagnostic models, achieving reasonable accuracy and helping to highlight the potential of computational methods in AD diagnosis.

2.1.2 Image Processing Methods

Traditional image processing methods were also pivotal in the early stages of Alzheimer's disease research. Techniques such as thresholding, edge detection, and region-based methods were employed to analyze neuroimaging data, particularly MRI and CT scans [1]. Thresholding involved setting intensity thresholds to segment brain images, highlighting regions of interest such as hippocampal atrophy, which is commonly associated with AD. Edge-based techniques focused on detecting boundaries and contours within the images, facilitating the identification of structural changes in the brain. Region-based methods aimed to segment images into meaningful regions, often using criteria such as intensity homogeneity or anatomical knowledge to delineate areas affected by the disease. These image processing methods were instrumental in extracting relevant features from neuroimaging data, which were subsequently used in classification models.

2.1.3 Limitations

While these earlier approaches made significant contributions to the field, they also had limitations. Machine learning techniques like SVM and LDA required extensive feature engineering and were often sensitive to the quality and quantity of input features. Additionally, traditional image processing methods were sometimes limited by their reliance on manual parameter tuning [2] and their susceptibility to noise and artifacts in the imaging data. Despite these challenges, the foundational work of binary classification and image processing paved the way for the development of more advanced models, including deep learning and multimodal approaches, which have since demonstrated superior performance in AD diagnosis.

2.2 Deep Learning Based Approaches

The landscape of Alzheimer's disease (AD) diagnosis has evolved significantly with the advent of advanced computational techniques, particularly in the realm of deep learning and multimodal data integration. Recent trends emphasize the use of sophisticated neural networks, leveraging the power of convolutional neural networks (CNNs), transformer-based models, and hybrid approaches to improve diagnostic accuracy and robustness.

2.2.1 Dominance of CNN Models

One of the most prominent trends in recent research is the application of deep learning methods, especially convolutional neural networks (CNNs) [26], to AD diagnosis. Unlike traditional machine learning techniques that require manual feature extraction, CNNs can automatically learn hierarchical features from raw imaging data. These models have been applied extensively to MRI and PET scans, providing state-of-the-art performance in distinguishing between AD, MCI, and CN individuals. For example, models like DenseNet [14] and ResNet [12] have been fine-tuned for the task, demonstrating substantial improvements in classification accuracy and computational efficiency.

2.2.2 Architecture Overview

2D CNN Based Architectures

The use of 2D Convolutional Neural Networks (CNNs) for classifying Alzheimer's Disease (AD) has been a prominent area of research. These methods typically involve extracting 2D slices from 3D MRI scans and using them as input for pre-trained or custom-designed CNN architectures.

Savaş [33] utilized pre-trained deep learning models, specifically VGG-16 and ResNet-50, for classifying AD stages through transfer learning. By fine-tuning these models on 2D MRI slices, the study leveraged learned features from the ImageNet dataset to enhance classification accuracy while reducing training time. Preprocessing ensured consistency and compatibility with the CNN input requirements.

- **VGG-16**: A deep CNN with 16 layers, pre-trained on ImageNet, served as a feature extractor. The final layers were replaced for AD classification.
- **ResNet-50**: A 50-layer residual network, also pre-trained on ImageNet, was fine-tuned similarly to VGG-16.

While effective, this approach might miss important 3D spatial information inherent in MRI scans. The fine-tuned models depend heavily on the specific dataset, potentially limiting their generalizability. Additionally, fine-tuning large pre-trained models still requires significant computational resources.

A CNN model [31] from scratch was developed to automate AD detection using 2D MRI slices. The model involved multiple convolutional layers for feature extraction, followed by ReLU activation functions and max-pooling layers to downsample feature maps and reduce computational complexity. Fully connected layers processed the

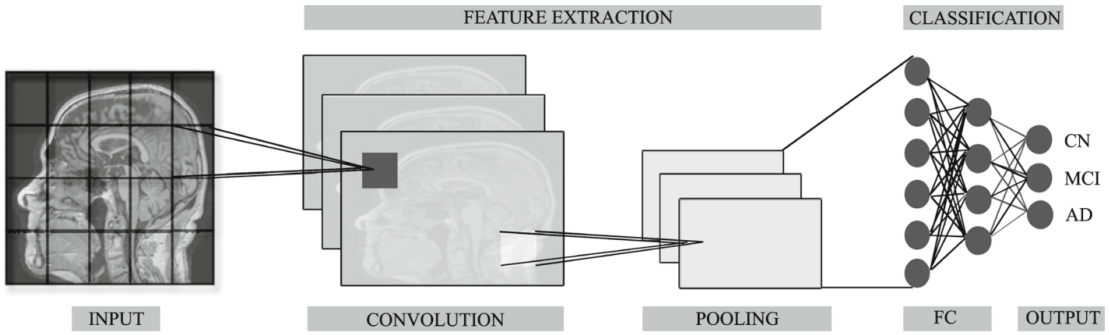


Fig. 2.1: A simple CNN architecture that extracts the features from the spatial information and then classify into different classes [33]

extracted features for classification. This model, however, requires a large amount of labeled data, which might not always be available. It also risks missing crucial 3D context and may overfit with limited datasets.

AlzheimerNet[34], a deep learning model for classifying AD stages from MRI images. AlzheimerNet incorporated convolutional blocks with batch normalization and ReLU activations, along with residual connections to mitigate vanishing gradients. Max-pooling layers downsampled feature maps, and fully connected layers enabled higher-level feature processing for classification.

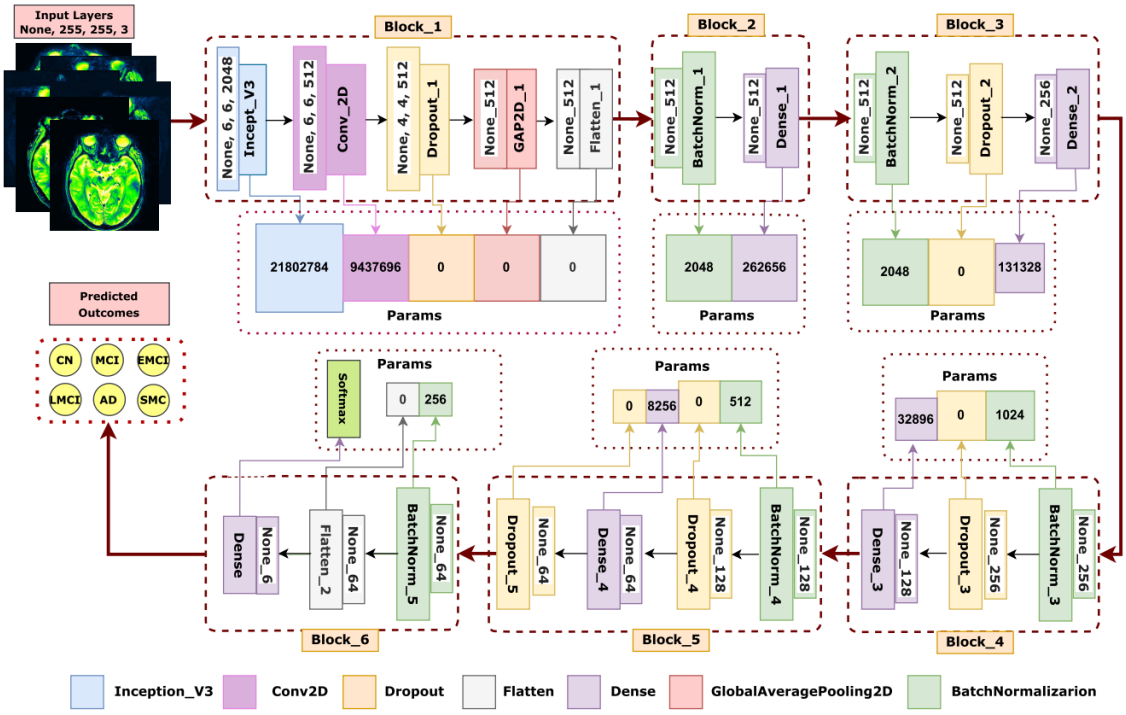


Fig. 2.2: The architecture of AlzheimerNet, a fine-tuned InceptionV3 [34]

AlzheimerNet's complex architecture, while powerful, may be computationally intensive for some clinical settings. The model's performance is dependent on the quality

and diversity of training data, and like other 2D CNN models, it may overlook important 3D spatial relationships critical for accurate AD diagnosis and staging.

Another study presents a novel deep-ensemble method [32] combining multiple convolutional neural network (CNN) architectures for robust and accurate classification of Alzheimer's Disease (AD) using MRI and fMRI data. The datasets include MRI and fMRI images of patients with varying degrees of dementia, including healthy controls, very mild AD, mild AD, and moderate AD. The CNN architectures selected, based on their performance in previous AD research and their size/precision ratios, include AlexNet, Inception-ResNet-v2, ResNet-50, ResNet-101, and GoogLeNet. Transfer learning was employed using CNNs pre-trained on the ImageNet dataset, which were fine-tuned on the AD datasets to adapt to the specific task of AD classification by retraining the networks' final layers while keeping the earlier layers frozen. Features were extracted from the penultimate fully connected layer (FC7) of AlexNet and the last layer of the ResNet and Inception architectures, providing a refined representation of the input images. An ensemble learning approach combined the predictions of the three best-performing networks (AlexNet, ResNet-101, and Inception-ResNet-v2) using a bagged trees model with an averaging strategy, enhancing the robustness and accuracy of classification. To prevent overfitting and improve model generalization, data augmentation techniques such as random rotation between -35° and 35°, random scaling in the x and y directions, and grey-scale preprocessing were applied to the training and validation sets.

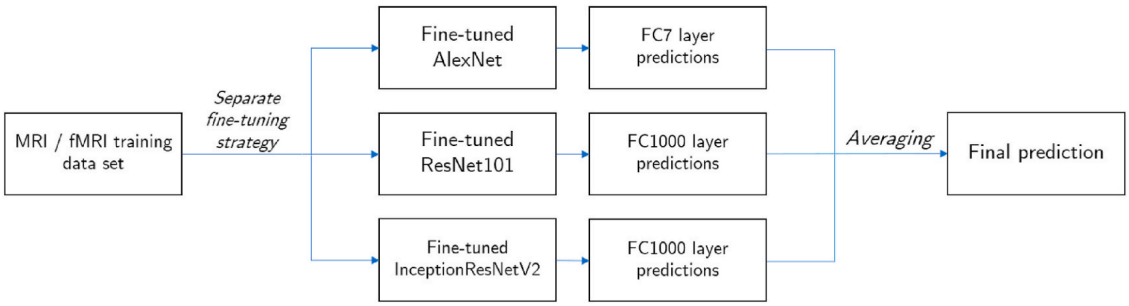


Fig. 2.3: Schematic representation of proposed ensemble model [21].

A robust classification model leveraged transfer learning with DenseNet and integrated within an embedded healthcare decision support system (DSS) [32]. Preprocessing steps included skull stripping, intensity normalization, and registration to a common reference space. DenseNet-121, pre-trained on the ImageNet dataset, was chosen for its dense connectivity patterns, which enhance feature reuse and mitigate the vanishing gradient problem. The pre-trained DenseNet was fine-tuned on the ADNI dataset by freezing initial layers, adding custom fully connected layers,

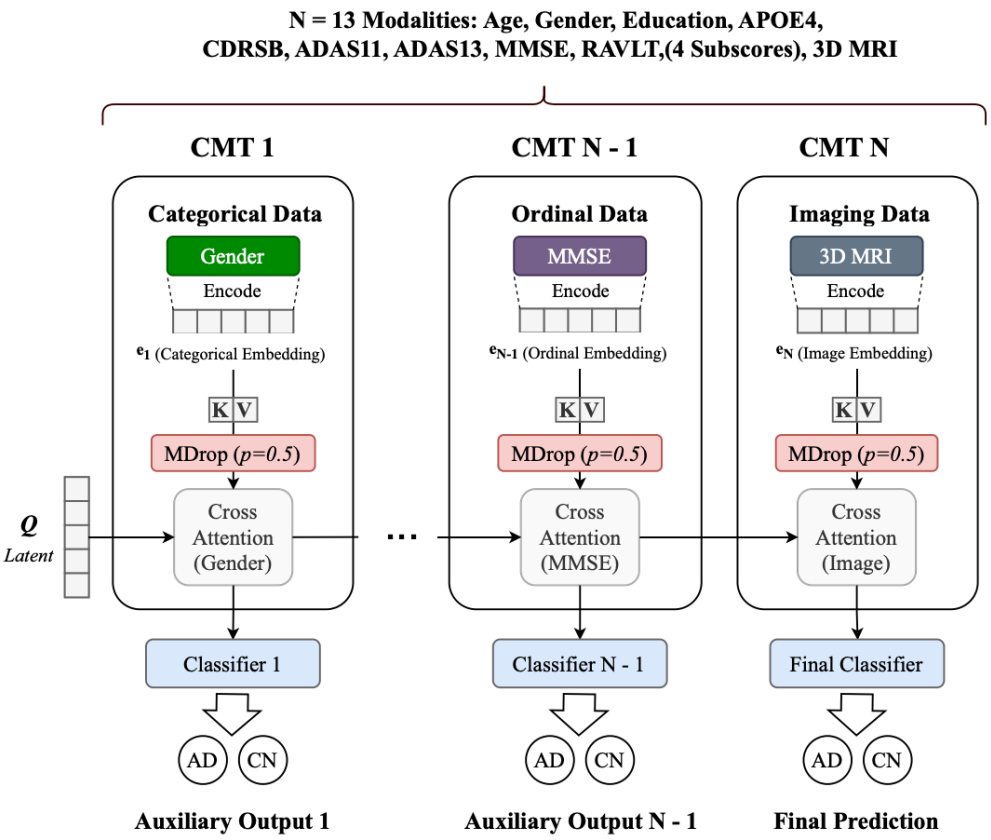


Fig. 2.4: Network architecture of 3MT [19].

and incorporating dropout regularization to prevent overfitting. The training process involved data augmentation techniques such as rotation, flipping, and scaling to improve model generalization, with optimal batch size and number of epochs determined experimentally. Despite high performance, limitations included dependency on the quality and availability of labeled MRI data, substantial computational resources required for DenseNet architecture, and the need for further validation to ensure generalization to different populations and imaging protocols beyond the ADNI dataset.

3D CNN Based Architectures

3D Convolutional Neural Networks (CNNs) are well-suited for medical imaging tasks involving volumetric data, such as MRI scans. These models process the entire 3D volume, capturing spatial relationships that might be missed by 2D approaches.

A cascaded multi-modal mixing transformer (CM3T) [19] framework was developed for AD classification using incomplete data. Their approach integrated 3D CNNs with transformers to handle the spatial complexity of MRI images. The CM3T model processed multi-modal neuroimaging data (structural MRI, fMRI, and PET scans) using

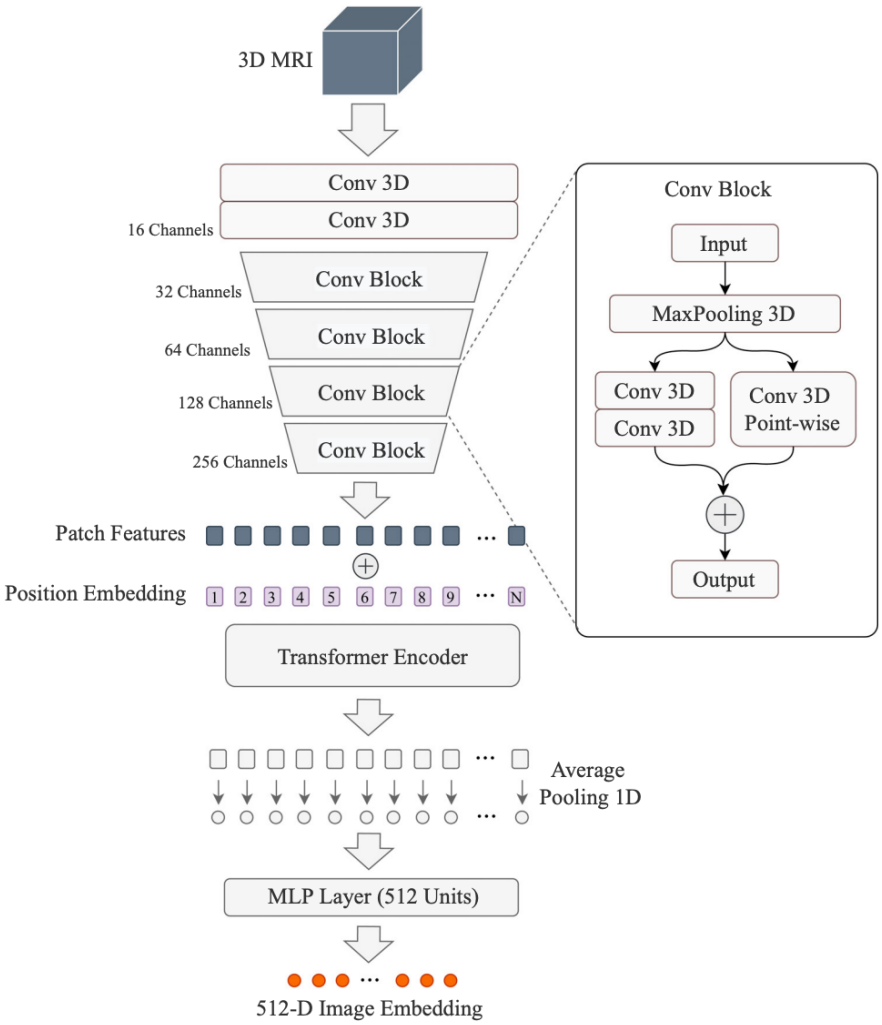


Fig. 2.5: Details of the CNN transformer for encoding 3D images [19].

dedicated transformer encoders for each modality. These encoders captured modality-specific representations and were fused using a multi-modal mixing module with cross-attention mechanisms (Fig. 2.4).

To handle incomplete data, the CM3T model employed a cascaded fusion strategy, progressively combining available modalities. The final multi-modal representation was fed into a classification head for AD stage prediction.

Helaly et al.[13] developed a 3D CNN for AD diagnosis using MRI volumes. Their model captured comprehensive anatomical features by processing the entire 3D brain volume, detecting subtle structural changes indicative of early-stage AD. The architecture consisted of multiple convolutional and pooling layers, followed by fully connected layers for classification. One of the primary strengths of this study lies in its high classification accuracy, particularly in detecting early-stage AD. By utilizing the entire 3D brain volume, the model demonstrated an exceptional ability to capture

detailed spatial information, potentially identifying subtle biomarkers that might be missed in 2D slice-based approaches. This comprehensive analysis of brain structure represents a significant advancement in the field of automated AD diagnosis. However, the study also faced several challenges. The 3D CNN approach requires substantial amounts of labeled 3D MRI data for effective training, which can be difficult and expensive to acquire in large quantities. Moreover, processing entire 3D volumes is computationally intensive, demanding significant processing power and memory resources. This could potentially limit the model's applicability in resource-constrained clinical settings. Another limitation noted was the model's propensity for overfitting, especially when trained on smaller datasets. This highlights the critical need for large, diverse datasets in developing robust 3D CNN models for AD diagnosis.

An interpretable deep learning framework [29] using 3D CNNs and MRI data was proposed for AD classification. Their model incorporated attention mechanisms to highlight critical brain regions contributing to classification decisions, enhancing interpretability and transparency. The 3D CNN processed full MRI volumes, capturing subtle patterns associated with AD. A significant strength of this study lies in its ability to achieve high classification accuracy while simultaneously providing interpretable results. The attention maps generated by the model offer valuable insights into the neural correlates of AD, potentially aiding clinicians in understanding the structural changes associated with the disease progression. This interpretability is crucial for building trust in AI-based diagnostic tools and could facilitate their integration into clinical workflows. However, the study also faced several challenges. Like many deep learning approaches in medical imaging, the model requires high-quality, diverse training data to perform optimally and generalize well to unseen cases. Acquiring such comprehensive datasets in the medical field remains a significant hurdle. Moreover, while the attention mechanisms enhance interpretability, the overall complexity of the model can still pose challenges in clinical settings. The intricate nature of deep learning models, even with interpretability features, may make it difficult for non-specialists to fully understand and trust the model's decisions. This highlights the ongoing challenge of balancing model sophistication with practical clinical applicability.

Utilizing the 3D CNNs Ebrahimi et al. [9] leveraged the rich spatial information in MRI volumes for AD detection. Their model employed multiple layers of 3D convolutions, followed by pooling and fully connected layers, to capture and process detailed anatomical features. A key strength of this study lies in its superior performance in AD detection compared to traditional 2D approaches. By leveraging the full 3D structure of the brain, the model demonstrated an enhanced ability to identify complex

spatial patterns associated with AD. This comprehensive volumetric analysis not only improved overall detection accuracy but also showed promise in enabling earlier and more precise AD diagnosis, a crucial factor in effective treatment and management of the disease. However, the study also faced significant challenges. The use of 3D CNNs necessitates extensive 3D MRI data for effective training, which can be difficult and costly to acquire in large quantities. Moreover, processing entire 3D volumes is computationally intensive, requiring substantial computational resources. This could potentially limit the model's applicability in resource-constrained clinical settings or research environments without access to high-performance computing facilities. Another critical consideration is the model's heavy dependence on the quality of the training data. The effectiveness of the 3D CNN in detecting AD is intrinsically linked to the comprehensiveness and accuracy of the MRI scans used for training. This underscores the importance of high-quality, diverse datasets in developing robust and generalizable models for AD detection.

Transformer Based Architectures

Recent advancements in deep learning have seen the emergence of transformer-based models, which have shown remarkable performance in various domains, including medical image analysis [35]. These models leverage the self-attention mechanism, which allows them to capture complex dependencies and interactions within the data, making them particularly well-suited for tasks that require detailed spatial and contextual understanding, such as Alzheimer's disease (AD) classification from MRI images.

The M3T model(Multi-Plane Multi-Slice Transformer) [16] combines transformers' self-attention capabilities with a multi-plane and multi-slice representation of 3D MRI volumes. This approach facilitates capturing complex spatial dependencies within the data. The model slices the 3D MRI volume into multiple 2D planes, which are further divided into smaller 2D slices treated as sequential inputs to the transformer model. The self-attention mechanism integrates information across multiple planes and slices, effectively reconstructing the 3D spatial context from the 2D representations. A key strength of this study lies in its significant improvement in classification accuracy over traditional CNN-based approaches. By leveraging the transformer's ability to model long-range dependencies, the M3T model demonstrates superior performance in capturing intricate spatial patterns and relationships within the brain structure. This enhanced capability is particularly crucial in AD classification, where subtle structural changes can be indicative of disease progression.

However, the study also faces several challenges. The complex architecture of the M3T model demands high computational resources, potentially limiting its application in resource-constrained environments. Additionally, like many deep learning approaches, the model's performance is heavily dependent on large, high-quality training datasets, which can be challenging to acquire in the medical imaging domain. Another significant consideration is the model's complexity, which can reduce its interpretability. While the model achieves high accuracy, the intricate nature of its decision-making process may be difficult for clinicians to understand and trust. This lack of transparency could potentially limit the model's acceptance and integration into clinical workflows, where interpretability is often crucial for decision-making and patient communication.

A novel approach combining pixel-level fusion techniques with vision transformers (ViTs) [25] was developed for early AD detection from MRI scans. ViTs effectively capture global relationships and dependencies within images [8]. Their method involves splitting high-resolution MRI images into smaller patches, embedding these into a sequence of tokens, and processing them with the transformer network. The pixel-level fusion technique ensures detailed and precise classification by integrating information from every part of the image. One of the primary strengths of this study lies in its superior performance in early AD detection. By leveraging ViTs, the model demonstrates an exceptional ability to capture long-range dependencies and global contextual information within brain MRI scans. This capability is particularly crucial in detecting subtle, early-stage indicators of AD that might be missed by traditional convolutional neural network (CNN) approaches, which typically focus on local features. However, the study also faces several significant challenges. The complex architecture of ViTs, combined with pixel-level fusion techniques, demands high computational power and memory resources. This requirement could potentially limit the model's applicability in resource-constrained clinical settings or research environments without access to high-performance computing facilities. Another notable limitation is the long training times associated with such complex model architectures. This not only impacts the development and iterative improvement of the model but also poses challenges for its adaptation to new datasets or different medical imaging modalities. Furthermore, a critical consideration for clinical applications is the difficulty in interpreting the model's decisions. While the model achieves high accuracy, the intricate nature of transformer architectures and pixel-level fusion makes it challenging to provide clear, interpretable explanations for its classifications. This lack of transparency could potentially hinder the model's acceptance and integration into clinical workflows, where interpretability is often crucial for decision-making and pa-

tient communication.

The Addformer model [18], which combines multiple transformer modules for multi-modal fusion of different MRI sequences (e.g., T1-weighted and T2-weighted images) for AD detection. Each MRI sequence is processed by a separate transformer module to capture modality-specific features, and the outputs are fused using additional transformer layers. This approach leverages the strengths of each MRI sequence and captures a comprehensive representation of the brain's structural characteristics. A key strength of this study lies in its enhanced robustness and accuracy in AD classification. By leveraging multiple MRI sequences, the Addformer model demonstrates an improved ability to detect subtle indicators of AD that might be more prominent in one modality than another. This multi-modal approach provides a more comprehensive view of brain structure and potential AD-related changes, potentially leading to more accurate and reliable diagnoses. Furthermore, the effective combination of multi-modal data represents a significant advancement in the field. The Addformer's architecture allows for the integration of complementary information from different MRI sequences, potentially capturing a wider range of AD biomarkers and structural changes associated with the disease progression. However, the study also faces several challenges. One notable limitation is the requirement for careful preprocessing and alignment of images from different modalities. This prerequisite adds complexity to the data preparation phase and may introduce potential sources of error if not handled meticulously. Another significant consideration is the increased computational requirements for both training and inference. The complex architecture of the Addformer, processing multiple MRI sequences through separate transformer modules, demands substantial computational resources. This could potentially limit the model's applicability in resource-constrained clinical settings or research environments without access to high-performance computing facilities. Moreover, the complexity of the Addformer model poses challenges in terms of interpretability. While the model achieves high accuracy, the intricate nature of its decision-making process, involving multiple transformer modules and fusion layers, may be difficult for clinicians to understand and trust. This lack of transparency could potentially hinder the model's acceptance and integration into clinical workflows, where interpretability is often crucial for decision-making and patient communication.

2.2.3 Multi-class Classification

While binary classification has been a common approach in Alzheimer's disease (AD) diagnosis, recent research has increasingly focused on multi-class classification to

better capture the spectrum of cognitive states associated with the disease. This approach distinguishes not only between Alzheimer's disease (AD) and cognitive normal (CN) individuals but also includes intermediate stages such as mild cognitive impairment (MCI). Multi-class classification is crucial for developing comprehensive diagnostic tools that can provide more nuanced assessments and support early intervention strategies.

2.2.4 Utilization of Single and Multi-modal Approaches

The diagnosis and classification of Alzheimer's disease (AD) have significantly advanced with the development of both single and multi-modal approaches. Each method offers unique benefits and, when combined, they provide a comprehensive and robust diagnostic framework.

Single-Modal Approaches

Single-modal approaches focus on analyzing data from a single imaging modality, typically MRI or PET scans. These methods have the advantage of being simpler and less resource-intensive compared to multi-modal approaches, making them more accessible in many clinical settings [21], [32], [33].

MRI-Based Approaches

- AlzheimerNet [34], a CNN-based model specifically designed to classify AD stages from functional brain changes observed in MRI images. The architecture utilized multiple convolutional layers to extract hierarchical features from 2D MRI slices, achieving notable accuracy in distinguishing between different stages of AD.
- A CNN model from scratch was designed to automate AD detection using MRI images [31]. By focusing on extensive data preprocessing, normalization, and augmentation techniques, their model demonstrated high robustness and accuracy in classification tasks, effectively handling the variability in MRI data.

PET-Based Approach: A 3D CNN framework [29] was developed to classify AD stages using PET images. Their model emphasized the interpretability of predictions, allowing clinicians to understand the decision-making process, thereby increasing the trust and usability of the model in clinical practice.

Multi-modal Approaches

Multi-modal approaches [30], [38] integrate information from multiple imaging modalities, such as MRI and PET, to leverage the complementary strengths of each. These methods provide a more comprehensive understanding of the brain's structure and function, enhancing diagnostic accuracy and robustness.

Fusion Techniques

Combining data from MRI and PET scans, multi-modal approaches use advanced deep learning architectures to integrate and analyze this diverse information.

- A proposed multimodal image fusion method [38] demonstrated superior performance in classifying Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Normal Control (NC) compared to single-modality approaches by leveraging both structural information from MRI and functional insights from Positron Emission Tomography (PET). The study utilized the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, including MRI and PET scans. Preprocessing involved skull stripping, intensity normalization, and registration for MRI, and intensity normalization and spatial alignment with MRI for PET. The core methodology included aligning PET images with MRI, extracting features from both using a modified ResNet, and combining features through a weighted averaging fusion strategy. The CNN architecture incorporated a modified ResNet and specialized 3D CNN models, including a U-shaped network with skip connections to extract multi-scale features. Data augmentation techniques like rotation, flipping, and scaling were applied to enhance variability. The model was optimized using the Adam optimizer with a learning rate scheduler and categorical cross-entropy loss function, reflecting a comprehensive approach to training and optimization. Despite high performance on the ADNI dataset, limitations include dependency on MRI and PET image quality and availability, additional computational complexity, and generalization challenges to other datasets and real-world scenarios.
- Addformer [18], a transformer-based model that fuses information from different MRI sequences. This model utilized multiple transformer modules to integrate data across various imaging modalities, enhancing robustness and accuracy in AD detection. The study demonstrated the potential of transformer-based models in leveraging multi-modal data for comprehensive AD classification.

Hybrid and Ensemble Methods

Hybrid approaches combine different deep learning models to utilize their respective strengths, often leading to superior performance in AD diagnosis.

- A cascaded multi-modal mixing transformer framework [19] that combines 3D CNNs with transformers. This hybrid method effectively handles the spatial complexity of MRI images, achieving robust classification even with incomplete data. The integration of CNNs and transformers highlighted the potential of hybrid models in improving diagnostic performance.
- A pixel-level fusion approach using vision transformers [25] for early AD detection. By processing MRI images at the pixel level, their model achieved detailed and precise classification, underscoring the advantages of transformers in handling high-resolution medical images. The study's multi-modal classification framework effectively distinguished between AD, MCI, and CN classes.

Transfer Learning and Domain Adaptation Transfer learning, where models pre-trained on large datasets are fine-tuned for specific tasks, has also been instrumental in multi-modal approaches. It allows the models to leverage learned features from extensive, general-purpose datasets, reducing the need for large amounts of domain-specific data.

While multi-modal approaches offer significant advantages, they also present challenges such as increased computational complexity, the need for synchronized multi-modal datasets, and the difficulty of integrating diverse data types. However, ongoing advancements in deep learning and computational power are addressing these challenges, paving the way for more efficient and effective multi-modal diagnostic tools.

In summary, the utilization of single and multi-modal approaches has greatly enriched the field of AD diagnosis. Single-modal methods, particularly those based on MRI and PET, provide valuable insights into the brain's structure and function. Multi-modal approaches, by integrating these insights, offer a more comprehensive and accurate diagnostic framework, ultimately enhancing the early detection and management of Alzheimer's disease.

2.2.5 Preprocessing Techniques

Preprocessing is crucial for the accuracy and efficiency of Convolutional Neural Network (CNN)-based models in MRI image analysis. Recent trends and popular techniques include:

- **Intensity Normalization**: Standardizes image intensities to aid feature learning. Common methods include Z-score normalization (mean of zero, standard deviation of one) and Min-Max scaling (fixed range, typically [0, 1] or [-1, 1]).
- **Bias Field Correction**: Corrects intensity inhomogeneities using algorithms like N4ITK to enhance image consistency [40].
- **Skull Stripping**: Removes non-brain tissues to focus on brain structures, improving classification accuracy. Common tools are Brain Extraction Tool (BET) from FSL [37], BrainSuite [36], and FreeSurfer [11].
- **Spatial Normalization**: Aligns MRI images to a common reference space, such as the MNI space, using affine or non-linear transformations to account for anatomical variability [6].
- **Segmentation**: Divides MRI images into tissue types (e.g., gray matter, white matter, cerebrospinal fluid) to isolate relevant brain structures. Tools include SPM [3] and FSL [37].
- **Data Augmentation**: Increases training dataset size and model robustness through random transformations such as rotation, translation, scaling, flipping, and adding Gaussian noise.
- **Smoothing**: Reduces noise and enhances signal-to-noise ratio using techniques like Gaussian smoothing to better delineate brain structures.
- **Patch Extraction**: Divides 3D MRI volumes into smaller patches for input to the CNN, reducing computational load and focusing on local features.
- **Histogram Equalization**: Enhances contrast by redistributing intensity values, improving visibility of brain structures for better feature learning.
- **Deep Learning-based Preprocessing**: Utilizes autoencoders and Generative Adversarial Networks (GANs) for denoising, normalizing, and enhancing MRI images in a data-driven manner.
- **Multimodal Image Fusion**: Combines different MRI modalities (e.g., T1-weighted, T2-weighted) to provide richer information for better classification performance through alignment and integration.
- **Domain Adaptation Techniques**: Addresses variations between training and testing data using techniques like adversarial domain adaptation to improve generalization by making learned features invariant to differences in scanners or protocols.

2.3 Summary and Limitations of Existing Architectures

In order to provide a concise comparison and highlight the key contributions of some of the studies which are aligned with our proposed methodology, we have summarized their core methodologies, model architectures, datasets used, and performance metrics in Table 2.1. This table offers a clear overview of the advancements and limitations observed in these studies, facilitating a better understanding of the current state of Alzheimer's disease classification using deep learning approaches.

In order to provide a concise comparison and highlight the key contributions of the representative studies analyzed in detail, we have summarized their core methodologies, results, and comments in Table 2.1. This table offers a clear overview of the advancements and limitations observed in these studies, facilitating a better understanding of the current state of Alzheimer's disease classification using deep learning approaches.

In summary, the reviewed literature underscores the effectiveness of advanced deep learning models, including 2D CNNs, 3D CNNs, and transformer-based classifiers, in Alzheimer's disease detection. However, each of these architectures has inherent limitations:

- **2D CNNs**: These models analyze individual 2D slices of MRI images, often failing to capture the interdependencies among slices. This slice-by-slice approach can lead to a loss of crucial 3D spatial information, which is vital for accurate AD diagnosis. Additionally, 2D CNNs can be computationally expensive when processing multiple slices separately.
- **3D CNNs**: While 3D CNNs are capable of analyzing volumetric data, they suffer from the "Curse of Dimensionality." The high number of parameters in these models makes them prone to overfitting, especially with limited training data. The non-convex nature of neural networks further complicates the learning process, reducing the chances of finding optimal parameters.
- **Transformer-based Classifiers**: These models leverage self-attention mechanisms to capture long-range dependencies in data. However, transformers require large amounts of data and computational resources for training, making them less suitable for datasets with limited samples. Additionally, they can be sensitive to the quality of input data and preprocessing techniques.

Table 2.1: Summary of Representative Works on Alzheimer's Disease Classification

Reference	Core Methodology	Results	Comments
Loddo et al. [21]	• Ensemble of multiple CNNs • Pre-trained on ImageNet • Simple average function for ensemble model (by averaging the top 3)	• Binary class accuracy: 99.29% • Dataset: ADNI	• Introduction of ensemble learning • Significant improvement in accuracy
Song et al. [38]	• Multimodal image fusion (MRI + PET) to create "GM-PET" images. • 3D CNN with U-Net-like architecture	• Binary class accuracy: 94.11% • Multi-class accuracy: 74.54% • Dataset: ADNI	• Highlights benefit of multimodal data • Need for additional data for optimization
Qiu et al. [30]	• Hybrid approach (CNN + CatBoost) • Integration of imaging and non-imaging data • Trained CNN model on MRI data to compute cognitive scores.	• Multi-class test accuracy: $(87.9 \pm 1.3)\%$	• Multiple independent datasets • Omitted CNN architecture details for MRI model • Extensive validation
Saleh et al. [32]	• Transfer learning with Densenet • Data augmentation	• Multi-class training accuracy: 96.05% • Multi-class testing accuracy: 90.01% • Dataset: Kaggle	• Dataset seems to contain less challenging data than ADNI • Indications of overfitting

Chapter 3

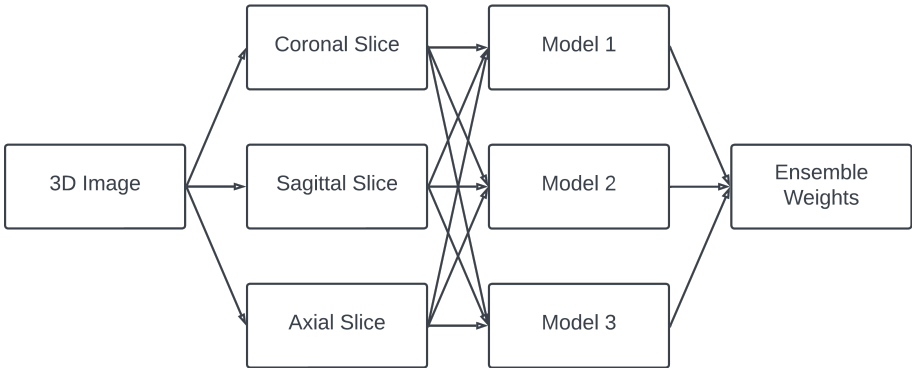
Proposed Methodology

3.1 Architecture Overview

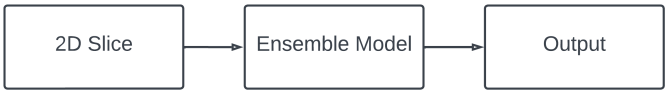
In this study, we employ an ensemble learning approach to enhance the classification performance of 3D MRI images for Alzheimer's Disease. The ensemble is composed of three distinct models, each specifically trained to process one of the three primary anatomical planes: coronal, sagittal, and axial. This approach, which we refer to as 'plane-wise ensemble', is designed to leverage the unique structural information inherent in each imaging plane. By integrating the outputs from models trained on these different planes, the approach aims to provide a more comprehensive and accurate classification than any single model could achieve alone. Fig. 3.1 provides a visual representation of this ensemble learning framework, highlighting the workflow and integration of the different models.

The rationale behind this plane-wise ensemble approach lies in the fact that different anatomical planes can reveal complementary aspects of brain structure and pathology. The coronal plane captures the frontal and posterior regions, the sagittal plane provides a lateral view, and the axial plane offers a top-down perspective. Each plane emphasizes different anatomical features, which, when combined, enhance the overall diagnostic accuracy.

This method is particularly advantageous in the context of 3D MRI image analysis, where the complexity and variability of brain structures necessitate a robust and multifaceted approach. By utilizing an ensemble of models, each attuned to specific structural information, our approach mitigates the limitations of individual models and capitalizes on the strengths of each perspective.

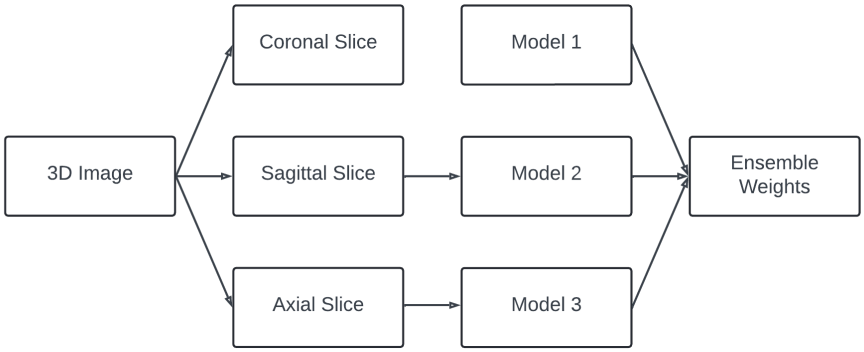


(a) Training (Forward Pass)

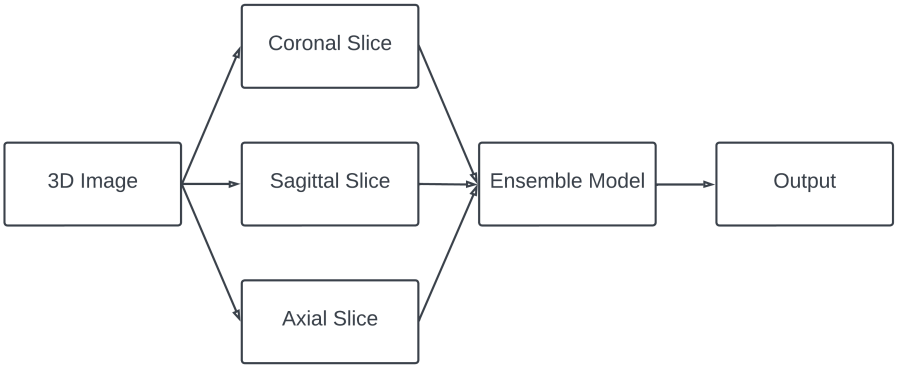


(b) Evaluation

(c) Regular Ensemble Model



(d) Training (Forward Pass)



(e) Evaluation

(f) Plane-wise Ensemble Technique

Fig. 3.1: Diagram comparing a regular ensemble model (c) with our proposed plane-wise ensemble technique (f)

3.2 Model Architecture

The three specialized models are trained on distinct MRI slice orientations to enhance Alzheimer's Disease classification. The Coronal Model focuses on coronal slices, capturing frontal to posterior brain structures such as the lateral ventricles and frontal lobes, which aids in identifying features unique to this plane. The Sagittal Model uses sagittal slices to highlight midline structures like the corpus callosum and brainstem, essential for detecting lateralized features and asymmetries. The Axial Model, trained on axial slices, captures horizontal structures including the cerebral cortex and basal ganglia, improving the detection of cortical thickness and basal ganglia configuration changes. Each model's specialization in its respective plane enhances its ability to identify Alzheimer's Disease-related patterns.

This segmentation of training data by plane is intended to enhance the models' ability to understand and interpret the slice-specific characteristics. By training models on specific imaging planes, each model can capture intricate details and features unique to its respective plane. This targeted approach minimizes internal confusion and maximizes the ability to identify subtle abnormalities, leading to the improvement of overall ensemble performance.

Each model's outputs are then integrated to form an ensemble, capitalizing on the strengths of each plane-specific model. The coronal, sagittal, and axial models collectively contribute to a more comprehensive analysis, with each model providing insights from different anatomical perspectives. This ensemble approach ensures that the final classification leverages the diverse and complementary information obtained from all three planes, resulting in a more robust and accurate diagnosis.

3.3 Model Integration

In the traditional ensemble model, when a single slice is provided as input, the models must first determine the anatomical plane to which the slice belongs before proceeding with classification. This approach effectively transforms the original 3-class classification problem into a more complex 9-class classification problem, as the model must now account for three planes for each class.

Unlike these traditional ensembles that evaluate individual 2D slices, our plane-wise ensemble processes entire 3D volumes, thereby providing a more comprehensive context for accurate classification. During the evaluation phase of the ensemble, an entire 3D MRI volume is used as input. This approach leverages the full spatial context of the

MRI data, capturing inter-slice relationships and volumetric features that are critical for accurate disease classification.

The 3D MRI volume is automatically sliced into the three primary anatomical planes: coronal, sagittal, and axial. Each set of slices corresponding to these planes is then fed into their respective specialized models within the ensemble. The coronal slices are processed by the coronal model, sagittal slices by the sagittal model, and axial slices by the axial model. This ensures that each model can apply its specialized knowledge to the appropriate set of slices, enhancing the detection of plane-specific features.

Following the classification of the slices by their respective models, the ensemble integrates the predictions by calculating a weighted average of the probabilities obtained from each model. These weights are not arbitrarily assigned; rather, they are learned through a separate training phase designed to optimize the combination of model outputs. This weighted averaging allows the ensemble to balance the contributions of each model according to their performance and relevance to the final classification.

By combining the strengths of the coronal, sagittal, and axial models, the plane-wise ensemble provides an accurate classification of the 3D MRI volume. This integration ensures that the diverse and complementary information from each anatomical plane is utilized effectively, leading to a more robust and reliable diagnosis.

3.4 Projector Functions

Projector functions play a critical role in the proposed plane-wise ensemble technique by transforming high-dimensional 3D MRI images into 2D inputs suitable for Convolutional Neural Network (CNN) models.

A *projector function* is a mathematical function that maps three-dimensional (3D) volumetric data into a series of two-dimensional (2D) images. Formally, let $V \subseteq \mathbb{R}^3$ represent the domain of 3D volumetric data. A projector function P can be defined as follows:

$$P : V \to (\mathbb{R}^2)^n \tag{3.1}$$

where $(\mathbb{R}^2)^n$ denotes an n-tuple of elements in $\mathbb{R}^2$, representing an ordered sequence of 2D images. For any point $\mathbf{v} \in V$, the projector function P maps $\mathbf{v}$ to a sequence of 2D images $(\mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_n)$ such that $\mathbf{i}_j \in \mathbb{R}^2$ for $j = 1, 2, \ldots, n$.

In the context of MRI image classification, these functions are used to decompose a

3D MRI scan into three sets of 2D planes: axial, coronal, and sagittal. This decomposition allows for the application of 2D CNN models, which are less computationally intensive than their 3D counterparts.

Projector functions are impactful for several reasons:

- **Reducing Computational Complexity:** Handling 3D MRI data directly with 3D CNNs can be computationally prohibitive due to the high dimensionality of the data. By projecting the 3D images into 2D planes, the projector functions significantly reduce the computational complexity, making it feasible to train and deploy CNN models on standard hardware.
- **Leveraging the Strengths of 2D CNNs:** 2D CNNs are well-established and widely used in image classification tasks, benefiting from extensive research and optimization. Projector functions enable the use of these robust 2D CNN architectures by converting the 3D MRI data into a format that these models can process effectively.
- **Maintaining Essential Structural Information:** Despite reducing the data dimensionality, projector functions try to ensure that the essential structural information of the brain is retained. By extracting slices in three orthogonal planes, the functions capture comprehensive views of the brain's anatomy, which is crucial for accurate disease diagnosis.
- **Enhancing Classification Accuracy and Efficiency:** The use of projector functions, in conjunction with an ensemble of CNN models trained on different planes, enhances the overall classification accuracy and efficiency. This approach allows the models to focus on specific aspects of the brain's structure, leveraging the strengths of each plane orientation to improve diagnostic performance.

Projector functions are a pivotal component of the proposed plane-wise ensemble technique for MRI image classification. By transforming 3D MRI data into 2D slices, they facilitate the use of efficient and accurate 2D CNN models, thereby addressing the computational challenges associated with 3D MRI data. This innovative approach enhances the performance and reliability of Alzheimer's disease diagnosis, contributing significantly to the field of medical imaging.

3.4.1 Midplane Projector

This function selects the middle slice of the 3D volume.

A 3D MRI scan can be considered as a stack of 2D images (slices) layered on top of each other. The midplane projector picks the slice that is exactly in the middle of this stack. This slice is considered representative of the entire volume and often captures a central view of the brain's structure.

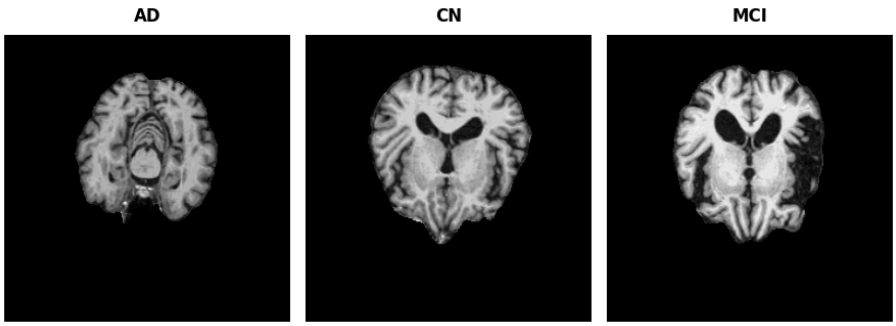


Fig. 3.2: Midplane Projections: Axial Plane

3.4.2 Average Projector

This function calculates the average of all slices in the 3D volume.

The average projector combines information from all slices by taking the average value for each pixel across the stack of slices. The resulting 2D image represents an averaged view, where each pixel's value is the mean of the corresponding pixels in the original 3D volume. This method aims to capture the overall intensity patterns in the brain.

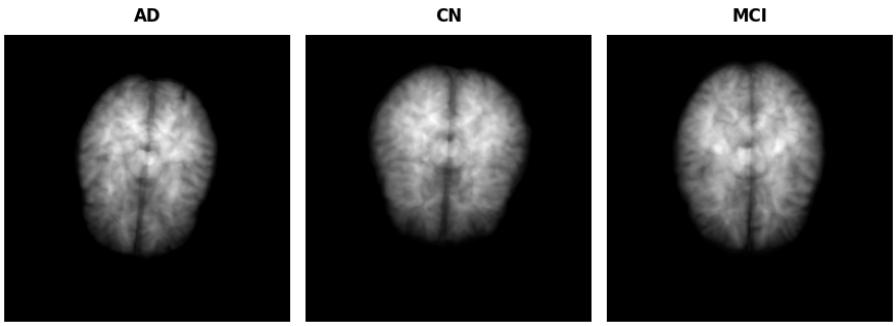


Fig. 3.3: Average Projections: Axial Plane

3.4.3 Max Variance Projector

This function selects the slice with the highest variance.

Variance measures how much the pixel values differ from the mean value within a slice. The max variance projector identifies the slice where the pixel values vary the most, indicating a high level of structural detail and differences.

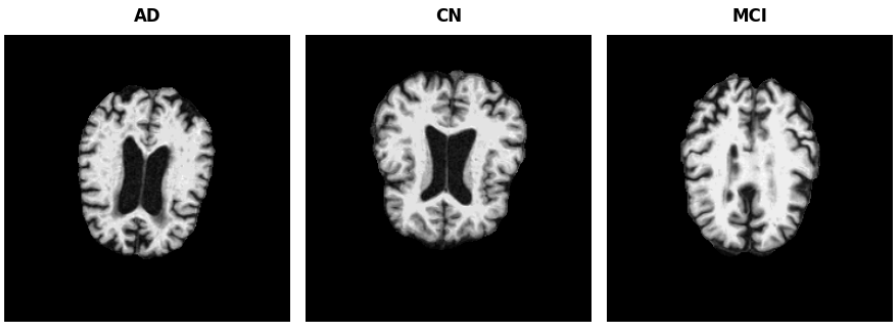


Fig. 3.4: Max Variance Projections: Axial Plane

3.4.4 Variance Weighted Average Projector

This function creates a weighted average of all slices, giving more importance to slices with higher variance.

Instead of treating all slices equally, the variance weighted average projector assigns greater weight to slices with higher variance, meaning those with more detail and differences. It then computes a weighted average, where each slice contributes to the final 2D image based on its variance. This approach aims to enhance the overall detail and information content in the projected image by emphasizing the most informative slices. Importantly, slices with very low variance, such as those that are fully black and contain no information, are assigned a weight of zero. This means they do not contribute to the final average, ensuring that only the slices with meaningful information are used in the projection.

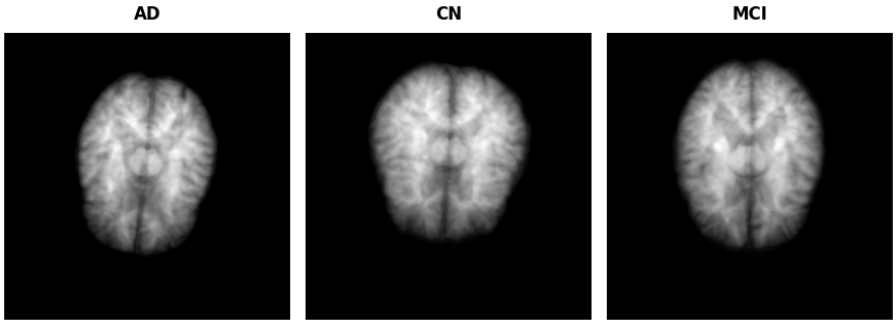


Fig. 3.5: Variance Weighted Average Projections: Axial Plane

3.4.5 Linear Learnable (LL) Projector

This function learns the optimal weights to assign to each slice in the 3D volume through a training process and then creates a weighted average of all slices using those weights.

Instead of manually defining the weights based on variance or other criteria, the LL

projector uses a machine learning model to determine how much weight each slice should contribute to the final 2D image. The working of this model can be seen as similar to an encoder-decoder model. The LL projector acts as the encoder, learning to compress and represent the 3D volume into a 2D plane through learned weights. The classifier that uses the 2D projection is analogous to the decoder, interpreting the 2D representation to make predictions.

During training, both the encoder and decoder update their weights based on the training samples and backpropagation. While evaluating novel images, the LL projector uses the learnt weights to project the 3D volume into 2D images. And then they are input to the planewise models to make the prediction.

Chapter 4

Results and Discussion

In this chapter, we have discussed the dataset, experimental setup and analysis process, emphasizing the advancements achieved in classification accuracy. It begins with a detailed explanation of dataset selection, preprocessing steps, and the methodology for splitting the data into training, validation, and test sets. Following this, we evaluate the performance of several widely-used convolutional neural network architectures, including AlexNet, VGG-16, and ResNet-50, using a comprehensive set of classification metrics. The chapter then introduces our novel plane-wise ensemble approach, explaining its implementation and demonstrating its advantages over traditional ensemble models. Through detailed performance metrics and comparative analysis, we demonstrate the effectiveness of our proposed method in improving classification accuracy across the three primary classes. The chapter concludes with a discussion of key evaluation metrics, providing a comprehensive view of model performance.

4.1 Dataset

The dataset utilized for our experiments was sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, which is widely recognized as a benchmark dataset for Alzheimer's Disease classification. The ADNI dataset includes a variety of imaging data, clinical information, and biomarkers collected from participants over multiple visits. This rich dataset is crucial for developing and validating models aimed at detecting and classifying Alzheimer's Disease.

As detailed in Table 4.1, the dataset comprises three primary classes: Alzheimer's Disease (AD), Cognitive Normal (CN), and Mild Cognitive Impairment (MCI). Each

Table 4.1: Dataset Distribution

Class	Number of Samples
AD (Alzheimer's Disease)	602
CN (Cognitive Normal)	998
MCI (Mild Cognitive Impairment)	1832

class includes a significant number of samples, with the MCI class having the largest representation.

One important aspect of the ADNI dataset is that it includes multiple imaging data points for individual patients collected across different visits. To ensure the integrity of the experimental results and prevent data leakage, we grouped all images from the same patient together. During the process of splitting the data into training, validation, and test sets, all images from a single patient are kept within the same split. This approach ensures that the models are evaluated on entirely unseen patients, thus providing a more realistic assessment of their generalizability and performance.

This careful handling of the dataset helps in maintaining the robustness of the training and evaluation process, ensuring that the models are not inadvertently trained on data that could appear in the test set. This practice is crucial for developing reliable models for Alzheimer's Disease classification, as it closely mirrors real-world scenarios where models must generalize well to new patients.

4.2 Experimental Setup

4.2.1 Data Preparation

Data Preprocessing

As illustrated in Fig. 4.1, the acquired MRI images underwent skull-stripping using the Freesurfer software. Freesurfer plays a vital role in preparing structural MRI data for Alzheimer's disease (AD) classification, offering a comprehensive pipeline to extract relevant features from brain images. The following steps outline the preprocessing process:

- **Intensity Normalization**: The intensity values of the input MRI scans are normalized to ensure consistency across different acquisitions. This step corrects variations in signal intensity caused by scanner differences and acquisition protocols.
- **Denoising Mechanisms**: Freesurfer incorporates denoising mechanisms to

reduce noise and enhance image quality. This includes the use of non-local means denoising, which preserves important anatomical details while effectively removing noise. This step is crucial for improving the accuracy of subsequent image processing tasks.

- **Skull Stripping and Semantic Segmentation**: Freesurfer employs a unified approach for skull stripping and semantic segmentation. By leveraging probabilistic atlases and machine learning algorithms, it accurately delineates brain structures while removing non-brain tissues such as skull, scalp, and dura mater. This step is crucial for eliminating extraneous structures and focusing on relevant brain regions for AD classification.
- **Tissue Segmentation**: Freesurfer performs tissue segmentation to classify voxels in the brain into different tissue types, including gray matter, white matter, and cerebrospinal fluid (CSF). This segmentation provides valuable information for subsequent analyses and feature extraction.
- **Cortical Surface Reconstruction**: Freesurfer reconstructs the cortical surface of the brain from MRI data, identifying the pial surface (outer boundary of the cortex) and the white matter surface (inner boundary of the cortex). Accurate cortical surface reconstruction is essential for capturing cortical morphological changes associated with AD.
- **Parcellation and Labeling**: Automated parcellation of the cerebral cortex into distinct anatomical regions is performed, enabling detailed analysis of cortical morphology and regional differences. Additionally, subcortical segmentation is carried out to delineate structures such as the hippocampus and basal ganglia, which are implicated in AD pathology.

By employing this preprocessing pipeline, Freesurfer prepares structural MRI data for AD classification studies, extracting relevant features and facilitating accurate analysis of brain morphometry and pathology.

Train-Test Split

The dataset underwent a partitioning process into train, validation (val), and test sets following an 8:1:1 ratio. To create the 5 folds for validation, the process involved independently generating each fold by randomly selecting a subset of samples from the dataset, ensuring that the classes are balanced within each subset. Each fold was produced using a 'sampling with replacement' approach, ensuring that the folds are independent of each other.

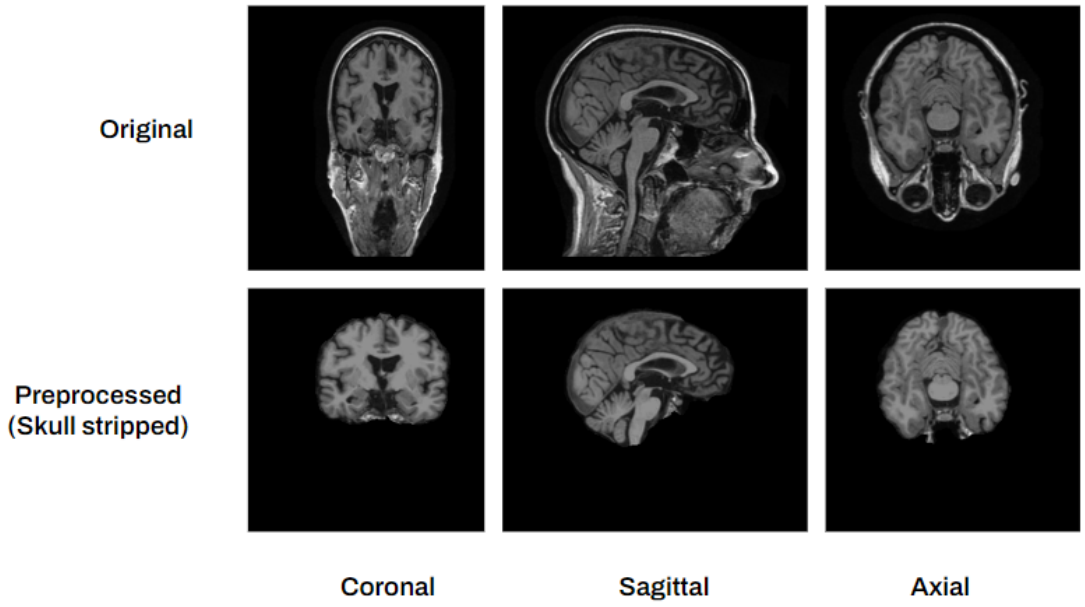


Fig. 4.1: Skull-stripping: The process of removing non-brain tissues from MRI images using semantic segmentation

The process of generating each fold involved the following steps:

- Randomly shuffle the dataset to ensure a fair distribution of samples.
- From the shuffled dataset, randomly select a subset such that the class distribution is balanced.
- Assign 80% of this subset to the training set, 10% to the validation set, and 10% to the test set.
- Repeat this process five times, independently, to create five separate folds.

To prevent data leakage, rigorous precautions were taken to ensure that multiple images from the same patient did not inadvertently end up across different splits during the random split generation. This meticulous approach was crucial for maintaining the integrity of the evaluation process. The distribution of data across these subsets is delineated in Table 4.2, providing transparency regarding the allocation of samples for training and evaluation.

4.2.2 Hyper-parameter Settings

The plane-wise ensemble model was trained using the following hyperparameters:

- **Batch Size:** 16
  The batch size determines the number of samples processed before the model's

Table 4.2: Data Split Distribution

Fold	Set	AD Samples	MCI Samples	CN Samples
1	Train	482	484	484
	Val	54	52	51
	Test	57	53	52
2	Train	480	484	482
	Val	51	47	49
	Test	61	58	56
3	Train	474	474	477
	Val	66	64	62
	Test	53	53	49
4	Train	474	477	473
	Val	63	61	62
	Test	57	54	57
5	Train	495	496	495
	Val	49	47	47
	Test	56	55	55

internal parameters are updated. A batch size of 16 strikes a balance between computational efficiency and the stability of the gradient descent process, enabling effective learning without overwhelming the memory capacity of the training hardware.

- **Maximum Number of Epochs:** 100
  An epoch refers to one complete pass through the entire training dataset. Setting the maximum number of epochs to 100 allows the model sufficient iterations to learn from the data while preventing excessive training time and overfitting.

- **Early Stopping:** Implemented to prevent overfitting, with a patience threshold set at 10 epochs
  Early stopping is a regularization technique used to terminate training when the model's performance on a validation set stops improving. The patience parameter of 10 epochs means that training will halt if there is no improvement in the validation loss for 10 consecutive epochs, thereby avoiding overfitting and reducing unnecessary computation.

- **Initial Learning Rate:** 0.001
  The learning rate controls the step size at each iteration while moving toward a minimum of the loss function. An initial learning rate of 0.001 is chosen as it is small enough to ensure stable convergence and large enough to expedite the learning process.

- **Scheduler Step Size:** 7 epochs
  The learning rate scheduler reduces the learning rate by a factor (gamma) every 7 epochs. This step size ensures periodic adjustments to the learning rate, facilitating finer learning adjustments as training progresses.
- **Gamma (Scheduler Factor):** 0.1
  The gamma parameter is the factor by which the learning rate is reduced. A gamma of 0.1 means the learning rate is multiplied by 0.1 every 7 epochs, allowing the model to fine-tune its weights with smaller learning rates in later stages of training for better accuracy.
- **Loss Function:** Cross-entropy loss
  Cross-entropy loss is employed due to its effectiveness in multiclass classification tasks. It measures the performance of the classification model whose output is a probability value between 0 and 1, helping to quantify the difference between predicted probabilities and the actual class labels.
- **Optimizer:** Adam
  Adam optimizer is used as it is the most popular default choice of optimizer in deep learning. It dynamically adjusts learning rates for individual parameters based on gradient magnitudes, smoothing convergence and reducing oscillations. Incorporating momentum, it aids in navigating loss landscapes, helping to escape shallow local optima. These features mitigate issues like jittering and local optima, enhancing training stability and speed.
- **Ensemble Weights Training:**
  The weights of the ensemble were fine-tuned separately for 30 epochs using the same hyperparameter settings. After training the individual models, the ensemble weights are optimized to combine the outputs of the models effectively. This separate training for 30 epochs with consistent hyperparameters ensures that the ensemble can leverage the strengths of each model and improve overall prediction accuracy.

4.3 Quantitative Evaluation

4.3.1 Evaluation Metrics

In the context of evaluating machine learning models, especially in classification tasks, several metrics are employed to gauge the performance of the model. These metrics offer insights into different aspects of the model's predictive capabilities.

This section delves into the specifics of accuracy, precision, recall, F1 score, and AUC-ROC, with a particular emphasis on their application in multi-class classification scenarios.

Accuracy

Accuracy is the simplest and most intuitive metric, representing the proportion of correctly classified instances out of the total instances. It is calculated as follows:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \tag{4.1}$$

In a multi-class classification setting, accuracy alone may not provide a complete picture, especially if the class distribution is imbalanced. For example, if one class dominates the dataset, a model that always predicts the majority class could still achieve high accuracy but would perform poorly on the minority classes.

Precision

Precision measures the proportion of true positive predictions among all positive predictions made by the model. It is an important metric when the cost of false positives is high. For multi-class classification, precision is calculated for each class individually:

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \tag{4.2}$$

where TP_i and FP_i are the true positives and false positives for class i, respectively. The overall precision for the model can be obtained by averaging the precision values for each class (macro-averaging) or by weighting them by the number of instances in each class (weighted averaging).

Recall

Recall, also known as sensitivity or true positive rate, measures the proportion of true positive predictions among all actual positive instances. It is crucial when the cost of false negatives is high. For multi-class classification, recall is calculated for each class as follows:

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \tag{4.3}$$

where TP_i and FN_i are the true positives and false negatives for class i, respectively. Similar to precision, overall recall can be obtained through macro-averaging or weighted averaging.

F1 Score

The F1 score is a metric that combines precision and recall into a single number by calculating their harmonic mean. It is particularly useful when dealing with imbalanced datasets, where one class may be significantly more frequent than others. In such cases, relying solely on accuracy can be misleading, as a model might perform well overall by favoring the majority class but poorly on the minority class.

The F1 score is also known as the Dice Score or Dice Coefficient in the context of certain applications like image segmentation. It is calculated for each class in a multi-class classification problem using the following formula:

$$\text{F1 Score}_i = \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{4.4}$$

The harmonic mean of two numbers, a and b, is given by:

$$H(a,b) = \frac{2}{\frac{1}{a} + \frac{1}{b}} = \frac{2ab}{a+b} \tag{4.5}$$

One of the key properties of the harmonic mean is that it tends to be closer to the smaller of the two numbers. This is beneficial in the context of the F1 score because it penalizes models that have a significant imbalance between precision and recall. In other words, if a model has high precision but low recall, or vice versa, the F1 score will be closer to the lower value, highlighting the model's deficiency.

AUC-ROC

The Receiver Operating Characteristic (ROC) curve is a graphical representation of a model's diagnostic ability. It plots the true positive rate (recall) against the false positive rate (1 - specificity) at various threshold settings. The Area Under the ROC Curve (AUC-ROC) summarizes the performance of the model across all thresholds:

$$\text{AUC-ROC} = \int_0^1 \text{ROC curve}(x)dx \tag{4.6}$$

For multi-class classification, a common approach is to compute the ROC curve and AUC for each class against all other classes (one-vs-rest) and then average the results.

- **True Positive Rate (TPR) or Recall**: This is the y-axis of the ROC curve and represents the proportion of actual positives correctly identified by the model.
- **False Positive Rate (FPR)**: This is the x-axis of the ROC curve and represents the proportion of actual negatives incorrectly identified as positives by the model.

A model with a high AUC-ROC value (closer to 1) indicates better performance, as it suggests that the model has a good measure of separability between the classes.

While the confusion matrix provides information about the actual counts of true positives, false positives, true negatives, and false negatives, it is dependent on a specific threshold. In contrast, the AUC-ROC offers a threshold-independent evaluation, summarizing the model's performance across all possible thresholds. This makes the AUC-ROC a more robust and comprehensive metric for assessing model performance, especially when comparing models.

In summary, each metric provides unique insights into different aspects of model performance. Accuracy offers a general overview, while precision and recall provide deeper insights into the model's behavior with respect to false positives and false negatives. The F1 score balances precision and recall, and the AUC-ROC provides a comprehensive evaluation across all thresholds. Together, these metrics form a holistic view of the model's performance in multi-class classification scenarios.

4.3.2 Baseline Models

We employed common 2D CNN models pretrained on the ImageNet [7] dataset as well as a traditional ensemble model, where each model was trained on the entirety of MRI slices, and predictions were combined through probability averaging during evaluation.

AlexNet

Table 4.3 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for AlexNet.

Table 4.3: Performance Analysis for AlexNet

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	60.494	AD	74.000	64.912	69.159	80.033
		MCI	45.946	65.385	53.968	61.626
		CN	71.053	50.943	59.341	71.733
2	60.000	AD	70.833	55.738	62.385	74.590
		MCI	46.667	62.500	53.435	66.161
		CN	69.231	62.069	65.455	78.912
3	59.355	AD	86.111	58.491	69.663	77.839
		MCI	43.210	71.429	53.846	66.712
		CN	68.421	49.057	57.143	71.162
4	55.357	AD	61.702	50.877	55.769	64.738
		MCI	43.750	61.404	51.095	62.162
		CN	70.732	53.704	61.053	70.110
5	60.843	AD	70.370	67.857	69.091	75.292
		MCI	45.161	50.909	47.863	58.624
		CN	70.000	63.636	66.667	71.122

VGG-16

Table 4.4 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for VGG-16.

Table 4.4: Performance Analysis for VGG-16

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	61.111	AD	79.070	59.649	68.000	79.866
		MCI	47.297	67.308	55.556	69.073
		CN	66.667	56.604	61.224	68.929
2	62.857	AD	76.471	63.934	69.643	74.863
		MCI	48.718	67.857	56.716	66.221
		CN	71.739	56.897	63.462	73.637
3	61.935	AD	88.889	60.377	71.910	72.660
		MCI	46.575	69.388	55.738	67.077
		CN	65.217	56.604	60.606	73.825
4	66.667	AD	80.000	56.140	65.979	79.501
		MCI	53.846	73.684	62.222	73.921
		CN	76.000	70.370	73.077	80.815
5	61.446	AD	70.833	60.714	65.385	75.633
		MCI	46.875	54.545	50.420	60.164
		CN	70.370	69.091	69.725	80.655

ResNet-50

Table 4.5 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for ResNet-50.

Table 4.5: Performance Analysis for ResNet-50

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	72.840	AD	75.000	78.947	76.923	84.712
		MCI	63.158	69.231	66.055	76.914
		CN	82.222	69.811	75.510	83.469
2	73.714	AD	80.769	68.852	74.336	79.062
		MCI	62.121	73.214	67.213	71.924
		CN	80.702	79.310	80.000	85.927
3	72.903	AD	69.091	71.698	70.370	74.769
		MCI	72.549	75.510	74.000	77.801
		CN	77.551	71.698	74.510	78.376
4	69.643	AD	68.254	75.439	71.667	78.125
		MCI	66.038	61.404	63.636	71.946
		CN	75.000	72.222	73.585	80.539
5	69.277	AD	68.966	71.429	70.175	79.140
		MCI	66.667	72.727	69.565	77.821
		CN	72.917	63.636	67.961	74.382

DenseNet-169

Table 4.6 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for DenseNet-169.

Table 4.6: Performance Analysis for DenseNet-169

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	77.778	AD	83.636	80.702	82.143	87.168
		MCI	69.231	69.231	69.231	79.528
		CN	80.000	83.019	81.481	88.783
2	74.857	AD	78.182	70.492	74.138	81.622
		MCI	67.143	83.929	74.603	83.283
		CN	82.000	70.690	75.926	81.359
3	74.839	AD	80.769	79.245	80.000	85.775
		MCI	65.000	79.592	71.560	80.400
		CN	81.395	66.038	72.917	84.462
4	75.000	AD	79.245	73.684	76.364	82.646
		MCI	68.852	73.684	71.186	78.694
		CN	77.778	77.778	77.778	86.160
5	73.494	AD	80.357	80.357	80.357	86.055
		MCI	62.687	76.364	68.852	77.674
		CN	81.395	63.636	71.429	78.870

Vision Transformer B16

Table 4.7 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Vision Transformer B16.

Table 4.7: Performance Analysis for Vision Transformer B16

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	64.815	AD	72.222	68.421	70.270	82.256
		MCI	56.667	65.385	60.714	70.472
		CN	66.667	60.377	63.366	70.573
2	60.571	AD	75.000	59.016	66.055	77.380
		MCI	51.351	67.857	58.462	70.273
		CN	60.377	55.172	57.658	66.858
3	70.968	AD	74.468	66.038	70.000	78.117
		MCI	64.912	75.510	69.811	78.283
		CN	74.510	71.698	73.077	79.967
4	61.310	AD	66.667	70.175	68.376	74.696
		MCI	53.125	59.649	56.198	67.520
		CN	65.909	53.704	59.184	70.744
5	65.663	AD	66.038	62.500	64.220	73.198
		MCI	60.000	65.455	62.609	73.726
		CN	71.698	69.091	70.370	79.230

Traditional Ensemble Model (ResNet-50, AlexNet, VGG-16)

The traditional ensemble model combines the strengths of three well-established Convolutional Neural Networks (CNNs): ResNet-50, AlexNet, and VGG-16. In this ensemble approach, each model is independently trained on the same dataset, and their predictions are aggregated to make the final decision. Table 4.8 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Traditional Ensemble Model (ResNet-50, AlexNet, VGG-16).

Table 4.8: Performance Analysis for Traditional Ensemble Model (ResNet-50, AlexNet, VGG-16)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	74.691	AD	84.211	84.211	84.211	87.347
		MCI	62.500	67.308	64.815	70.359
		CN	77.551	71.698	74.510	80.898
2	73.714	AD	80.392	67.213	73.214	77.858
		MCI	65.079	73.214	68.908	75.040
		CN	77.049	81.034	78.992	85.427
3	81.935	AD	90.909	75.472	82.474	86.632
		MCI	74.000	75.510	74.747	80.485
		CN	81.967	94.340	87.719	90.972
4	74.405	AD	81.481	77.193	79.279	83.750
		MCI	71.429	61.404	66.038	71.189
		CN	70.769	85.185	77.311	81.453
5	73.494	AD	75.806	83.929	79.661	82.976
		MCI	64.516	72.727	68.376	73.095
		CN	83.333	63.636	72.165	76.717

4.3.3 Plane-wise Ensemble Models

The proposed plane-wise ensemble, with models trained specifically on coronal, sagittal, and axial slices, demonstrated superior performance compared to traditional ensemble models and standalone 2D CNNs. The holistic approach of combining specialized models using a weighted average of their predictions yielded improved accuracy in the classification of 3D MRI images.

Triple ResNet (using Midplane Projector)

Triple ResNet model utilizes a plane-wise ensemble technique, involving three separate ResNet-50 models trained on coronal, sagittal, and axial planes of MRI images. Each ResNet-50 model is trained independently on 2D slices from one of these planes, capturing unique anatomical features specific to that orientation. During evaluation, the entire 3D MRI volume is processed by a midplane projector function, which extracts the central slice from each of the three planes. These slices are then fed into their respective ResNet-50 models. The predictions from the three models are subsequently combined to make the final classification decision. Table 4.9 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Triple ResNet (Midplane Projector).

Table 4.9: Performance Analysis for Triple ResNet (using Midplane Projector)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	80.247	AD	85.455	82.456	83.929	89.323
		MCI	70.175	76.923	73.394	81.189
		CN	86.000	81.132	83.495	88.264
2	68.000	AD	88.000	72.131	79.279	83.937
		MCI	52.055	67.857	58.915	69.703
		CN	71.154	63.793	67.273	74.536
3	70.323	AD	81.395	66.038	72.917	78.635
		MCI	56.522	79.592	66.102	76.627
		CN	81.395	66.038	72.917	78.487
4	77.381	AD	85.965	85.965	85.965	89.584
		MCI	67.797	70.175	68.966	76.403
		CN	78.846	75.926	77.358	83.138
5	72.892	AD	81.132	76.786	78.899	82.240
		MCI	63.492	72.727	67.797	73.661
		CN	76.000	69.091	72.381	77.952

Triple DenseNet (using Midplane Projector)

Similar to Triple ResNet, Triple DenseNet model works similarly by using three separate DenseNet-169 models, each trained on separate planes of MRI images. Table 4.10

shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Triple DenseNet (Midplane Projector).

Table 4.10: Performance Analysis for Triple DenseNet (using Midplane Projector)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	68.519	AD	79.167	66.667	72.381	79.382
		MCI	56.897	63.462	60.000	72.867
		CN	71.429	75.472	73.394	80.474
2	75.429	AD	90.741	80.328	85.217	88.165
		MCI	62.903	69.643	66.102	77.101
		CN	74.576	75.862	75.214	84.468
3	76.774	AD	91.111	77.358	83.673	85.553
		MCI	61.765	85.714	71.795	79.842
		CN	85.714	67.925	75.789	81.021
4	81.548	AD	90.385	82.456	86.239	89.663
		MCI	76.786	75.439	76.106	82.741
		CN	78.333	87.037	82.456	89.214
5	76.506	AD	81.818	80.357	81.081	87.565
		MCI	67.797	72.727	70.175	79.279
		CN	80.769	76.364	78.505	84.734

Triple ResNet (using Average Projector)

Table 4.11 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Triple ResNet (Average Projector).

Table 4.11: Performance Analysis for Triple ResNet (using Average Projector)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	82.716	AD	94.118	84.211	88.889	90.443
		MCI	70.312	86.538	77.586	84.965
		CN	87.234	77.358	82.000	87.052
2	83.429	AD	89.831	86.885	88.333	93.198
		MCI	72.727	85.714	78.689	87.410
		CN	90.000	77.586	83.333	87.902
3	80.000	AD	97.727	81.132	88.660	90.474
		MCI	63.889	93.878	76.033	85.637
		CN	89.744	66.038	76.087	82.556
4	83.333	AD	92.157	82.456	87.037	91.355
		MCI	76.190	84.211	80.000	88.620
		CN	83.333	83.333	83.333	87.817
5	86.747	AD	94.118	85.714	89.720	91.315
		MCI	78.689	87.273	82.759	89.091
		CN	88.889	87.273	88.073	94.529

Triple DenseNet (using Average Projector)

Table 4.12 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Triple DenseNet (Average Projector).

Table 4.12: Performance Analysis for Triple DenseNet (using Average Projector)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	83.951	AD	87.719	87.719	87.719	90.426
		MCI	76.271	86.538	81.081	88.392
		CN	89.130	77.358	82.828	89.060
2	84.000	AD	94.231	80.328	86.726	89.617
		MCI	70.667	94.643	80.916	89.931
		CN	93.750	77.586	84.906	87.371
3	85.806	AD	89.130	77.358	82.828	90.233
		MCI	75.000	91.837	82.569	91.721
		CN	95.918	88.679	92.157	94.802
4	82.738	AD	85.965	85.965	85.965	87.514
		MCI	73.438	82.456	77.686	81.618
		CN	91.489	79.630	85.149	87.021
5	81.928	AD	87.755	76.786	81.905	88.166
		MCI	68.000	92.727	78.462	86.798
		CN	100.000	76.364	86.598	90.975

Triple ResNet (using Max Variance Projector)

Table 4.13 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Triple ResNet (Max Variance Projector).

Table 4.13: Performance Analysis for Triple ResNet (using Max Variance Projector)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	79.012	AD	83.333	78.947	81.081	81.437
		MCI	67.742	80.769	73.684	81.521
		CN	89.130	77.358	82.828	87.900
2	73.143	AD	80.000	72.131	75.862	77.869
		MCI	59.420	73.214	65.600	70.408
		CN	84.314	74.138	78.899	84.807
3	75.484	AD	97.561	75.472	85.106	89.271
		MCI	61.017	73.469	66.667	75.144
		CN	74.545	77.358	75.926	80.818
4	75.000	AD	84.906	78.947	81.818	83.404
		MCI	63.235	75.439	68.800	76.324
		CN	80.851	70.370	75.248	80.702
5	74.699	AD	80.000	78.571	79.279	87.744
		MCI	62.712	67.273	64.912	77.183
		CN	82.692	78.182	80.374	84.521

Triple DenseNet (using Max Variance Projector)

Table 4.14 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Triple DenseNet (Max Variance Projector).

Table 4.14: Performance Analysis for Triple DenseNet (using Max Variance Projector)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	82.099	AD	88.333	92.982	90.598	92.648
		MCI	72.222	75.000	73.585	85.000
		CN	85.417	77.358	81.188	87.277
2	72.000	AD	83.019	72.131	77.193	82.312
		MCI	57.353	69.643	62.903	70.888
		CN	79.630	74.138	76.786	79.458
3	69.032	AD	77.551	71.698	74.510	83.888
		MCI	52.830	57.143	54.902	63.188
		CN	77.358	77.358	77.358	81.835
4	75.595	AD	84.906	78.947	81.818	87.182
		MCI	64.062	71.930	67.769	77.146
		CN	80.392	75.926	78.095	82.732
5	71.687	AD	87.805	64.286	74.227	82.971
		MCI	56.944	74.545	64.567	71.188
		CN	79.245	76.364	77.778	79.640

Triple ResNet (using Variance Weighted Average Projector)

Table 4.15 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Triple ResNet (Variance Weighted Average Projector).

Table 4.15: Performance Analysis for Triple ResNet (using Variance Weighted Average Projector)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	84.568	AD	95.833	80.702	87.619	91.328
		MCI	70.588	92.308	80.000	90.175
		CN	93.478	81.132	86.869	88.852
2	88.571	AD	93.443	93.443	93.443	97.412
		MCI	78.788	92.857	85.246	92.077
		CN	95.833	79.310	86.792	87.563
3	89.032	AD	92.157	88.679	90.385	93.896
		MCI	81.481	89.796	85.437	90.027
		CN	94.000	88.679	91.262	94.229
4	85.119	AD	90.566	84.211	87.273	89.268
		MCI	74.286	91.228	81.890	85.728
		CN	95.556	79.630	86.869	85.981
5	82.530	AD	89.796	78.571	83.810	90.227
		MCI	71.642	87.273	78.689	85.242
		CN	90.000	81.818	85.714	89.287

Triple DenseNet (using Variance Weighted Average Projector)

Table 4.16 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Triple DenseNet (Variance Weighted Average Projector).

Table 4.16: Performance Analysis for Triple DenseNet (using Variance Weighted Average Projector)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	85.185	AD	91.071	89.474	90.265	92.231
		MCI	78.571	84.615	81.481	84.318
		CN	86.000	81.132	83.495	89.476
2	89.143	AD	96.364	86.885	91.379	92.954
		MCI	77.941	94.643	85.484	91.747
		CN	96.154	86.207	90.909	93.914
3	87.097	AD	90.566	90.566	90.566	92.194
		MCI	76.786	87.755	81.905	89.623
		CN	95.652	83.019	88.889	93.563
4	85.714	AD	92.000	80.702	85.981	87.751
		MCI	75.000	89.474	81.600	87.798
		CN	94.000	87.037	90.385	93.535
5	88.554	AD	94.340	89.286	91.743	95.357
		MCI	78.788	94.545	85.950	90.270
		CN	95.745	81.818	88.235	88.468

Triple ResNet (using LL Projector)

Table 4.17 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Triple ResNet (LL Projector).

Table 4.17: Performance Analysis for Triple ResNet (using LL Projector)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	83.951	AD	92.308	84.211	88.073	88.705
		MCI	71.429	86.538	78.261	84.388
		CN	91.489	81.132	86.000	87.883
2	82.286	AD	87.931	83.607	85.714	88.510
		MCI	72.414	75.000	73.684	79.862
		CN	86.441	87.931	87.179	87.179
3	87.097	AD	91.489	81.132	86.000	92.064
		MCI	75.410	93.878	83.636	93.685
		CN	97.872	86.792	92.000	95.838
4	85.714	AD	90.196	80.702	85.185	88.083
		MCI	77.419	84.211	80.672	85.696
		CN	90.909	92.593	91.743	94.851
5	88.554	AD	92.593	89.286	90.909	92.776
		MCI	81.034	85.455	83.186	85.946
		CN	92.593	90.909	91.743	94.824

Triple DenseNet (using LL Projector)

Table 4.18 shows the metrics for accuracy, classwise precision, recall, F1 score, and AUC-ROC for each of the 5 folds for Triple DenseNet (LL Projector).

Table 4.18: Performance Analysis for Triple DenseNet (using LL Projector)

Fold	Accuracy (%)	Class	Precision (%)	Recall (%)	F1 Score (%)	AUC-ROC (%)
1	88.272	AD	94.340	87.719	90.909	92.080
		MCI	78.125	96.154	86.207	89.650
		CN	95.556	81.132	87.755	89.043
2	84.000	AD	91.525	88.525	90.000	91.444
		MCI	72.727	85.714	78.689	88.205
		CN	90.000	77.586	83.333	88.609
3	82.581	AD	93.750	84.906	89.109	92.823
		MCI	69.643	79.592	74.286	81.767
		CN	86.275	83.019	84.615	88.161
4	79.762	AD	88.000	77.193	82.243	87.293
		MCI	68.493	87.719	76.923	84.953
		CN	88.889	74.074	80.808	86.972
5	87.349	AD	92.727	91.071	91.892	94.756
		MCI	79.661	85.455	82.456	84.848
		CN	90.385	85.455	87.850	87.666

4.3.4 Comparative Analysis and Discussion

The plane-wise ensemble approach emerged as a promising strategy, surpassing the performance of both traditional ensemble models and standalone Convolutional Neural Networks (CNNs). This novel methodology demonstrated the efficacy of tailoring models for different imaging planes, showcasing a nuanced understanding of the underlying data structure.

One of the notable strengths of the plane-wise ensemble was its ability to leverage reduced class complexity inherent in specialized models for individual imaging planes. By focusing on specific planes, each model could better capture the unique features and patterns present within them, leading to improved classification accuracy. Moreover, this targeted approach allowed for more efficient utilization of computational resources, as models were optimized for their respective domains.

Additionally, the incorporation of plane-specific information provided a valuable contextual framework for the classification task. By considering the orientation and characteristics of each imaging plane, the ensemble could effectively exploit domain-specific knowledge to enhance decision-making processes. This contextual understanding facilitated more accurate predictions, particularly in cases where subtle variations between planes significantly influenced diagnostic outcomes.

Furthermore, the utilization of the entire 3D image volume during evaluation conferred a distinct advantage to the plane-wise ensemble. By integrating information from multiple planes, the ensemble could capture spatial relationships and dependencies that may not be discernible when analyzing individual slices in isolation. This

holistic approach enabled a more comprehensive assessment of pathological features, leading to improved diagnostic accuracy and robustness.

Table 4.19 provides a comprehensive overview of the performance metrics for various models, highlighting the superiority of the plane-wise ensemble approach. Several observations can be made based on these results:

- **Challenges of Multi-Class Classification**: A notable trend observed across various models and splits is the consistently lower recall scores for the Mild Cognitive Impairment (MCI) class compared to Alzheimer's Disease (AD) and Cognitive Normal (CN) classes. This disparity underscores the inherent difficulty of multi-class classification tasks, particularly in distinguishing between subtle cognitive states. The minimal overlap between the AD and CN classes further accentuates the distinctiveness of these conditions, posing challenges for accurate classification in the presence of intermediary states such as MCI.
- **Impact of Plane-Wise Ensemble Technique**: Models utilizing the plane-wise ensemble technique not only achieves better accuaracy but also demonstrate lower variance in performance metrics across different splits compared to traditional ensemble models and standalone CNNs. This reduced variance signifies greater stability and consistency in classification outcomes, indicating that the plane-wise ensemble approach is less susceptible to fluctuations in data characteristics or training conditions. The ability of plane-wise ensembles to harness the strengths of specialized models for individual imaging planes contributes to this stability, as each model focuses on capturing relevant features within its domain, thereby reducing ambiguity and uncertainty in classification decisions.

Overall, the combination of reduced class complexity, inherent plane information, and holistic 3D evaluation underscored the effectiveness of this methodology in medical image analysis.

We can use the data from Table 4.19 to estimate the population mean accuracy. The sample mean accuracy and sample standard deviation of accuracy can be used to find a confidence interval of the population mean accuracy. For a small sample size (less than 30), we use the t-distribution to calculate confidence intervals. The t-distribution takes into account the sample size and provides more accurate estimates of the population mean when the sample size is small.

The confidence intervals in the table are calculated using the formula:

Table 4.19: Model Performance Comparison

Projector	Model	Accuracy (%)					Mean (%)	Std (%)
		Fold-1	Fold-2	Fold-3	Fold-4	Fold-5		
-	AlexNet	60.494	60.000	59.355	55.357	60.843	59.210	2.225
	VGG-16	61.111	62.857	61.935	66.667	61.446	62.803	2.258
	ResNet-50	72.840	73.714	72.903	69.643	69.277	71.675	2.056
	DenseNet-169	77.778	74.857	74.839	75.000	73.494	75.193	1.569
	Vision Transformer B16	64.815	60.571	70.968	61.310	65.663	64.665	4.146
	Ensemble (ResNet-50, AlexNet, VGG-16)	74.691	73.714	81.935	74.405	73.494	75.648	3.549
Midplane	Triple ResNet	80.247	68.000	70.323	77.381	72.892	73.768	5.023
	Triple DenseNet	68.519	75.429	76.774	81.548	76.506	75.755	4.681
Average	Triple ResNet	82.716	83.429	80.000	83.333	86.747	83.245	2.404
	Triple DenseNet	83.951	84.000	85.806	82.738	81.928	83.685	1.471
Max Variance	Triple ResNet	79.012	73.143	75.484	75.000	74.699	75.468	2.167
	Triple DenseNet	82.099	72.000	69.032	75.595	71.687	74.083	5.053
Variance Weighted Average	Triple ResNet	84.568	88.571	**89.032**	85.119	82.530	85.964	2.769
	Triple DenseNet	85.185	**89.143**	87.097	**85.714**	**88.554**	**87.139**	1.722
Linear Learnable	Triple ResNet	83.951	82.286	87.097	**85.714**	**88.554**	85.520	2.483
	Triple DenseNet	**88.272**	84.000	82.581	79.762	87.349	84.393	3.488

$$CI = \mu \pm t\left(\frac{s}{\sqrt{n}}\right) \tag{4.7}$$

where CI is the confidence interval, μ is the sample mean accuracy, s is the sample standard deviation, n is the sample size and t is the critical value from the t-distribution with $(n-1)$ degree of freedom corresponding to the desired confidence level and sample size.

In comparison to state-of-the-art models in Alzheimer's disease (AD) classification from neuroimaging data, our proposed model demonstrates competitive performance across various metrics. Table 4.21 provides a comparison between our model and representative state-of-the-art models.

Table 4.20: Confidence Intervals for Model Accuracy

Projector	Model	Accuracy (%)		Confidence Interval		
		Mean	Std	80% CI	90% CI	95% CI
-	AlexNet	59.210	2.225	(57.685, 60.735)	(57.089, 61.331)	(56.448, 61.972)
	VGG-16	62.803	2.258	(61.255, 64.351)	(60.650, 64.956)	(60.000, 65.606)
	ResNet-50	71.675	2.056	(70.265, 73.085)	(69.715, 73.635)	(69.123, 74.227)
	DenseNet-169	75.193	1.569	(74.117, 76.269)	(73.697, 76.689)	(73.245, 77.141)
	Vision Transformer	64.665	4.146	(61.823, 67.507)	(60.712, 68.618)	(59.518, 69.812)
	Ensemble (ResNet-50, AlexNet, VGG-16)	75.648	3.549	(73.215, 78.081)	(72.264, 79.032)	(71.242, 80.054)
Midplane	Triple ResNet	73.768	5.023	(70.324, 77.212)	(68.979, 78.557)	(67.532, 80.004)
	Triple DenseNet	75.755	4.681	(72.546, 78.964)	(71.292, 80.218)	(69.944, 81.566)
Average	Triple ResNet	83.245	2.404	(81.597, 84.893)	(80.953, 85.537)	(80.261, 86.229)
	Triple DenseNet	83.685	1.471	(82.677, 84.693)	(82.282, 85.088)	(81.859, 85.511)
Max Variance	Triple ResNet	75.468	2.167	(73.982, 76.954)	(73.402, 77.534)	(72.778, 78.158)
	Triple DenseNet	74.083	5.053	(70.619, 77.547)	(69.265, 78.901)	(67.810, 80.356)
Variance Weighted Average	Triple ResNet	85.964	2.769	(84.066, 87.862)	(83.324, 88.604)	(82.526, 89.402)
	Triple DenseNet	**87.139**	1.722	(85.958, 88.320)	(85.497, 88.781)	(85.001, 89.277)
Linear Learnable	Triple ResNet	85.520	2.483	(83.818, 87.222)	(83.153, 87.887)	(82.437, 88.603)
	Triple DenseNet	84.393	3.488	(82.002, 86.784)	(81.067, 87.719)	(80.063, 88.723)

Table 4.21: Comparison with State-of-the-art Models

Reference	**Method**	**Accuracy (%)**	
		Mean	**Std**
Song et al. [38]	3D CNN with U-Net like architecture	74.54	-
Qiu et al. [30]	Hybrid approach (CNN + CatBoost)	87.9	1.3
Saleh et al. [32]	Transfer learning with DenseNet	**90.01**	-
Ours	Triple DenseNet (Variance Weighted Average Projector)	**87.139**	1.722
Ours	Triple ResNet (Variance Weighted Average Projector)	85.964	2.769
Ours	Triple ResNet (LL Projector)	85.520	2.483
Ours	Triple DenseNet (LL Projector)	84.393	3.488

Chapter 5

Conclusion

In this work, we presented a novel CNN-based pipeline for Alzheimer's Disease classification using 3D MRI images through a Plane-wise Ensemble Technique. Our approach addresses the critical challenge of enhancing diagnostic accuracy while optimizing computational efficiency in the analysis of 3D MRI data for Alzheimer's Disease classification. This research was motivated by the increasing prevalence of Alzheimer's disease and the need for early, accurate diagnosis to improve patient outcomes and support the development of effective therapeutic strategies.

Utilizing the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, a benchmark in AD research, we worked with a comprehensive set of 3,432 MRI samples across three classes: Alzheimer's Disease (AD), Cognitive Normal (CN), and Mild Cognitive Impairment (MCI). Our methodology, which involves decomposing 3D data into axial, coronal, and sagittal planes and training specialized models for each, achieved significant improvements in diagnostic accuracy. This method effectively leverages the unique anatomical information present in each plane, demonstrating the power of multi-view integration in medical image analysis.

Our research has made several significant contributions to the field. We introduced an innovative ensemble technique that capitalizes on plane-specific information, offering a more comprehensive analysis of brain structure. This approach not only enhances classification accuracy but also provides a more nuanced understanding of the anatomical features associated with Alzheimer's Disease progression. Additionally, we developed efficient projector functions that streamline the transformation of 3D data into 2D representations, addressing a critical preprocessing challenge and making the model more feasible for clinical application even with limited computational resources. Furthermore, our extensive benchmarking of various CNN architectures

provides valuable insights into model performance for Alzheimer's Disease classification. Each of these contributions advances the field in meaningful ways, from improving analytical techniques to enhancing the practical applicability of our approach in clinical settings.

The implications of our findings extend beyond improved accuracy. Our research underscores the importance of incorporating detailed anatomical information during model training, offering a promising avenue for enhancing medical image diagnosis across various conditions. The Plane-wise Ensemble Technique not only improves classification outcomes but also provides a more comprehensive view of brain structure changes associated with AD.

Despite these advancements, we acknowledge certain limitations in our study. The need for further exploration into optimal model selection for each imaging plane and ensemble model optimization presents opportunities for future research. Additionally, while we've made strides in computational efficiency, further optimization may be necessary for widespread clinical adoption.

Looking ahead, several key areas warrant further investigation. Advanced model selection tailored to each imaging plane could potentially enhance the accuracy of individual classifiers. Optimization of the ensemble model through diverse architecture combinations may yield even more robust results. Furthermore, exploration of additional image preprocessing techniques, such as advanced registration methods, could further refine feature extraction and overall system performance.

As we continue to refine these techniques, the potential for improved early diagnosis and treatment of Alzheimer's Disease becomes increasingly promising. This work not only advances our understanding of CNN-based approaches in medical image analysis but also opens new avenues for enhancing diagnostic accuracy and efficiency across various neurodegenerative disorders. The impact of this research extends beyond the technical realm, offering hope for patients and their families by contributing to the development of more effective diagnostic tools and treatment strategies. By making advanced diagnostic capabilities more accessible, we aim to support global healthcare systems in managing the growing challenges posed by Alzheimer's Disease and related cognitive impairments.

References

[1] S. Afzal, M. Maqsood, U. Khan, *et al.*, "Alzheimer disease detection techniques and methods: A review," 2021.

[2] J. Arslan, H. Jamshed, and H. Qureshi, "Early detection and prevention of alzheimer's disease: Role of oxidative markers and natural antioxidants," *Frontiers in Aging Neuroscience*, vol. 12, 2020, ISSN: 1663-4365. DOI: 10.3389/fnagi.2020.00231. [Online]. Available: https://www.frontiersin.org/journals/aging-neuroscience/articles/10.3389/fnagi.2020.00231.

[3] J. Ashburner and K. J. Friston, "Voxel-based morphometry—the methods," *NeuroImage*, vol. 11, no. 6, pp. 805–821, 2000, ISSN: 1053-8119. DOI: https://doi.org/10.1006/nimg.2000.0582. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811900905822.

[4] M. Basso, J. Yang, L. Warren, *et al.*, "Volumetry of amygdala and hippocampus and memory performance in alzheimer's disease," *Psychiatry Research: Neuroimaging*, vol. 146, no. 3, pp. 251–261, 2006, ISSN: 0925-4927. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925492706000084.

[5] D. L. Beekly, E. M. Ramos, W. W. Lee, *et al.*, "The national alzheimer's coordinating center (nacc) database: The uniform data set," *Alzheimer Disease & Associated Disorders*, vol. 21, no. 3, pp. 249–258, 2007.

[6] M. Brett, I. S. Johnsrude, and A. M. Owen, "The problem of functional localization in the human brain," *Nature Reviews Neuroscience*, vol. 3, no. 3, pp. 243–249, 2002.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.

[8] A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, *An image is worth 16x16 words: Transformers for image recognition at scale*, 2021. arXiv: 2010.11929 [cs.CV]. [Online]. Available: https://arxiv.org/abs/2010.11929.

[9] A. Ebrahimi, S. Luo, and A. D. N. Initiative, "Convolutional neural networks for Alzheimer's disease detection on MRI images," *Journal of Medical Imaging*, vol. 8, no. 2, p. 024 503, 2021. DOI: 10.1117/1.JMI.8.2.024503. [Online]. Available: https://doi.org/10.1117/1.JMI.8.2.024503.

[10] K. A. Ellis, A. I. Bush, D. Darby, *et al.*, "The australian imaging, biomarkers and lifestyle (aibl) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease," *International Psychogeriatrics*, vol. 21, no. 4, pp. 672–687, 2009. DOI: 10.1017/S1041610209009405.

[11] B. Fischl, "Freesurfer," *NeuroImage*, vol. 62, no. 2, pp. 774–781, 2012, 20 YEARS OF fMRI, ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2012.01.021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811912000389.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[13] H. A. Helaly, M. Badawy, and A. Y. Haikal, "Deep learning approach for early detection of alzheimer's disease," *Cognitive computation*, vol. 14, no. 5, pp. 1711–1727, 2022.

[14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.

[15] C. R. Jack Jr., M. A. Bernstein, N. C. Fox, *et al.*, "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008. DOI: https://doi.org/10.1002/jmri.21049. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.21049. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.21049.

[16] J. Jang and D. Hwang, "M3t: Three-dimensional medical image classifier using multi-plane and multi-slice transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 20 718–20 729.

[17] J. Kim, M. Jeong, W. R. Stiles, and H. S. Choi, "Neuroimaging modalities in alzheimer's disease: Diagnosis and clinical features," *International Journal of Molecular Sciences*, vol. 23, no. 11, p. 6079, 2022. DOI: 10.3390/ijms23116079. [Online]. Available: https://doi.org/10.3390/ijms23116079.

[18] R. Kushol, A. Masoumzadeh, D. Huo, S. Kalra, and Y.-H. Yang, "Addformer: Alzheimer's disease detection from structural mri using fusion transformer," in

2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, pp. 1–5. DOI: 10.1109/ISBI52829.2022.9761421.

[19] L. Liu, S. Liu, L. Zhang, X. V. To, F. Nasrallah, and S. S. Chandra, "Cascaded multi-modal mixing transformers for alzheimer's disease classification with incomplete data," *NeuroImage*, vol. 277, p. 120 267, 2023, ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2023.120267. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811923004184.

[20] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of alzheimer's disease with deep learning," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, 2014, pp. 1015–1018. DOI: 10.1109/ISBI.2014.6868045.

[21] A. Loddo, S. Buttau, and C. Di Ruberto, "Deep learning based pipelines for alzheimer's disease diagnosis: A comparative study and a novel deep-ensemble method," *Computers in Biology and Medicine*, vol. 141, p. 105 032, 2022, ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2021.105032. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001048252100826X.

[22] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open Access Series of Imaging Studies: Longitudinal MRI Data in Nondemented and Demented Older Adults," *Journal of Cognitive Neuroscience*, vol. 22, no. 12, pp. 2677–2684, Dec. 2010, ISSN: 0898-929X. DOI: 10.1162/jocn.2009.21407. eprint: https://direct.mit.edu/jocn/article-pdf/22/12/2677/1940172/jocn.2009.21407.pdf. [Online]. Available: https://doi.org/10.1162/jocn.2009.21407.

[23] J. C. Morris, M. Storandt, J. P. Miller, *et al.*, "Mild Cognitive Impairment Represents Early-Stage Alzheimer Disease," *Archives of Neurology*, vol. 58, no. 3, pp. 397–405, Mar. 2001, ISSN: 0003-9942. DOI: 10.1001/archneur.58.3.397. eprint: https://jamanetwork.com/journals/jamaneurology/articlepdf/778838/noc00280.pdf. [Online]. Available: https://doi.org/10.1001/archneur.58.3.397.

[24] K. O'Shea and R. Nash, *An introduction to convolutional neural networks*, 2015. arXiv: 1511.08458 [cs.NE]. [Online]. Available: https://arxiv.org/abs/1511.08458.

[25] M. Odusami, R. Maskeliūnas, and R. Damaševičius, "Pixel-level fusion approach with vision transformer for early detection of alzheimer's disease," *Electronics*, vol. 12, no. 5, 2023, ISSN: 2079-9292. DOI: 10.3390/electronics12051218. [Online]. Available: https://www.mdpi.com/2079-9292/12/5/1218.

[26] V. Patil, M. Madgi, and A. Kiran, "Early prediction of alzheimer's disease using convolutional neural network: A review," *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, vol. 58, no. 1, Article 130, 2022. DOI: 10.1186/s41983-022-00515-w. [Online]. Available: https://ejnpn.springeropen.com/articles/10.1186/s41983-022-00515-w.

[27] A. Payan and G. Montana, "Predicting alzheimer's disease: A neuroimaging study with 3d convolutional neural networks," *CoRR*, vol. abs/1502.02506, 2015. arXiv: 1502.02506. [Online]. Available: http://arxiv.org/abs/1502.02506.

[28] R. D. Piovezan, D. Oliveira, N. Arias, D. Acosta, M. J. Prince, and C. P. Ferri, "Mortality rates and mortality risk factors in older adults with dementia from low-and middle-income countries: The 10/66 dementia research group population-based cohort study," *Journal of Alzheimer's Disease*, vol. 75, no. 2, pp. 581–593, 2020.

[29] S. Qiu, P. S. Joshi, M. I. Miller, *et al.*, "Development and validation of an interpretable deep learning framework for Alzheimer's disease classification," *Brain*, vol. 143, no. 6, pp. 1920–1933, May 2020, ISSN: 0006-8950. DOI: 10.1093/brain/awaa137. eprint: https://academic.oup.com/brain/article-pdf/143/6/1920/33446057/awaa137.pdf. [Online]. Available: https://doi.org/10.1093/brain/awaa137.

[30] S. Qiu, M. I. Miller, P. S. Joshi, *et al.*, "Multimodal deep learning for alzheimer's disease dementia assessment," *Nature communications*, vol. 13, no. 1, p. 3404, 2022.

[31] P. M. Raees and V. Thomas, "Automated detection of alzheimer's disease using deep learning in mri," in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1921, 2021, p. 012 024.

[32] A. W. Saleh, G. Gupta, S. B. Khan, N. A. Alkhaldi, and A. Verma, "An alzheimer's disease classification model using transfer learning densenet with embedded healthcare decision support system," *Decision Analytics Journal*, vol. 9, p. 100 348, 2023, ISSN: 2772-6622. DOI: https://doi.org/10.1016/j.dajour.2023.100348. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772662223001881.

[33] S. Savaş, "Detecting the stages of alzheimer's disease with pre-trained deep learning architectures," *Arabian Journal for Science and Engineering*, vol. 47, no. 2, pp. 2201–2218, 2022.

[34] F. J. M. Shamrat, S. Akter, S. Azam, *et al.*, "Alzheimernet: An effective deep learning based proposition for alzheimer's disease stages classification from

functional brain changes in magnetic resonance images," *IEEE Access*, vol. 11, pp. 16 376–16 395, 2023.

[35] F. Shamshad, S. Khan, S. W. Zamir, *et al.*, "Transformers in medical imaging: A survey," *Medical Image Analysis*, vol. 88, p. 102 802, 2023, ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2023.102802. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841523000634.

[36] D. W. Shattuck and R. M. Leahy, "Brainsuite: An automated cortical surface identification tool," *Medical Image Analysis*, vol. 6, no. 2, pp. 129–142, 2002, ISSN: 1361-8415. DOI: https://doi.org/10.1016/S1361-8415(02)00054-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841502000543.

[37] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, 2002. DOI: https://doi.org/10.1002/hbm.10062. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.10062. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.10062.

[38] J. Song, J. Zheng, P. Li, X. Lu, G. Zhu, and P. Shen, "An effective multimodal image fusion method using mri and pet for alzheimer's disease diagnosis," *Frontiers in Digital Health*, vol. 3, 2021, ISSN: 2673-253X. DOI: 10.3389/fdgth.2021.637386. [Online]. Available: https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2021.637386.

[39] S. Teipel, A. Drzezga, M. J. Grothe, *et al.*, "Multimodal imaging in alzheimer's disease: Validity and usefulness for early detection," *The Lancet Neurology*, vol. 14, no. 10, pp. 1037–1053, 2015, ISSN: 1474-4422. DOI: https://doi.org/10.1016/S1474-4422(15)00093-9. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474442215000939.

[40] N. J. Tustison, B. B. Avants, P. A. Cook, *et al.*, "N4itk: Improved n3 bias correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010. DOI: 10.1109/TMI.2010.2046908.

[41] Z. Zhao, J. H. Chuah, K. W. Lai, *et al.*, "Conventional machine learning and deep learning in alzheimer's disease diagnosis using neuroimaging: A review," *Frontiers in Computational Neuroscience*, vol. 17, 2023, ISSN: 1662-5188. DOI: 10.3389/fncom.2023.1038636. [Online]. Available: https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2023.1038636.